THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Préparée dans le cadre d'une cotutelle *entre l'Université Grenoble Alpes* et l'*Università degli Studi di Milano*

Spécialité : **CHIMIE - BIOLOGIE** et **CHIMIE**

Arrêté ministériel : le 6 janvier 2005 – 25 mai 2016

Présentée par

Rafael BERMEO MALO

Thèse dirigée par **Annabelle VARROT** et **Anna BERNARDI**

préparée au sein du **Centre de Recherche sur les Macromolécules Végétales** et du **Département de Chimie de l' l'Université de Milan**

dans **l'École Doctorale Chimie et Sciences du Vivant** et **l'École doctorale de Chimie de l'Université de Milan**

Conception, synthèse et évaluation de glycocomposés dirigés contre BC2L-C

Thèse soutenue publiquement le **18 Juin 2021**,
devant le jury composé de :

Dr. Joanna TIMMINS
Directrice de Recherche CNRS - UGA, Présidente
Prof. Luigi LAY
Université de Milan, Examinateur
Prof. Ulf NILSSON
Université de Lund, Examinateur
Prof. Cristina NATIVI
Université de Florence, Rapportrice
Prof. Alexander TITZ
Université de Saarland, Rapporteur

THESIS

To obtain the degree of

DOCTOR OF GRENOBLE ALPES UNIVERSITY

Prepared as part of a joint supervision between the *Grenoble Alpes University* and the *University of Milan*

Specialty: **CHEMISTRY - BIOLOGY** and **CHEMISTRY**

Ministerial decree: January 6, 2005 - May 25, 2016

Presented by

Rafael BERMEO MALO

Thesis supervised by **Annabelle VARROT** and **Anna BERNARDI**

prepared in the **Center for Research on Plant Macromolecules** and the **Chemistry Department of University of Milan**

in **the Doctoral School of Chemistry and Life Sciences** and **the Doctoral School of Chemistry of University of Milan**

Design, synthesis and evaluation of antagonists towards BC2L-C

Thesis publicly defended on **June 18 2021**,
Before the jury composed by:

Dr. Joanna TIMMINS
Research Director CNRS - UGA, Jury President
Prof. Luigi LAY
University of Milan, Examiner
Prof. Ulf NILSSON
University of Lund, Examiner
Prof. Cristina NATIVI
University of Florence, Referee
Prof. Alexander TITZ
University of Saarland, Referee

TESI DI DOTTORATO DI RICERCA IN CO-TUTELA

UNIVERSITÀ DEGLI STUDI DI MILANO

CORSO DI DOTTORATO in **CHIMICA**

Scuola di Dottorato in Chimica

Ciclo XXXIII

DIPARTIMENTO DI CHIMICA

UNIVERSITY GRENOBLE ALPES

DOCTORATE in **CHEMISTRY - BIOLOGY**

Doctoral School of Chemistry and Life Sciences

TESI DI DOTTORATO DI RICERCA:

Design, synthesis and evaluation of antagonists towards BC2L-C

Rafael BERMEO MALO

Matr. Nr. R12344

Tutors:

Prof. Anna Bernardi

Dr. Annabelle Varrot

Coordinator:

Prof. Emanuela Licandro

Anno accademico 2020/2021

We choose to do these things not because they are easy, but because they are hard.

John F. Kennedy

- Can one still be brave if he is afraid?

- That is the only time one can be brave.

George R. R. Martin

May the problems we face today become the lessons we teach tomorrow.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765581

Acknowledgements

Besides *Horizon 2020*, there is a long list of people without whom this project would have been impossible.

First and foremost is Annabelle Varrot, who has been an exceptional mentor and a splendid person: she has really made me into a better scientist and taught me a number of new things that wouldn't fit on this thesis. I can really say that I now know twice as much as I did at the beginning of this journey, and in no small part it's thanks to her. Always displaying patience and understanding, Annabelle has been a steady point of reference during these years. I cannot thank you enough!

Right next to her, Anna Bernardi has been an invaluable influence on my development as a chemist: the extent of scientific knowledge and experience she has can only be compared to the remarkable care she puts into sharing it with the next generation of chemists under her wing. Anna can always be trusted to push you towards excellence with the feeling she's got your back. Thank you for supporting and preparing me for successful future ventures so well!

Besides their essential contribution to this journey, these two mentors have managed to surround themselves (and thus, me) with exceptional people, who have been excellent scientific and personal influences: among many, many scientific models of excellence, I have to thank Anne Imberty, Serge Perez and Laura Belvisi, as well as other members of the PhD4GlycoDrug consortium who have, in one way or another, contributed to my scientific development and to the enrichment of this project. A special thanks goes to the coordinator of this European endeavour: Marko Anderluh, who has carried on with the consortium through excellent as well as difficult times, always making me feel grateful for being a part of such a well taken-cared-for team. Other good influences were found in Franck Fieschi and Luigi Lay, who joined me early on by contributing to my yearly thesis evaluation, and Olivier Renaudet, who joined soon afterwards, and had the most amazing willingness to support my project in uncertain times. I thank all of your for inspiring me and showing me that wonderful scientists are, without fail, wonderful people.

A very special thanks goes to the members of my thesis committee: Joanna Timmins, Cristina Nativi, Alexander Titz, Ulf Nilsson and Luigi Lay, who showed good will by taking interest in my work, and will take the time to read this thesis and join me for the defence.

Annabelle and Anna have also managed to establish wonderful teams. In Grenoble, I was warmly welcomed in 2018 by Oriane and Aurore, who guided me through the early days in the lab, and François, who shared my office and was always pleasant to hang out with. Naturally, the GBMS team isn't itself without Valou, Emilie and Olivier, who were almost as important as Annabelle in showing me the intricacies and workings of the GMBS group. Other members of the group that were less present were also always nice to meet occasionally. (There was also the occurrence of Dania, which will be discussed in section Dania).

Moving through time, the Bernardi group, along with the Gennari group, had the warmest of receptions for me, and really welcomed me to discover and enjoy the Italian culture and language, and improve my calcetto skills! Thank you to Marco, Lorenzo, Crescenzo, Luca, Saretta, and Albi and Giovanni for always being friendly and willing to help me and make me feel welcome in Italy. Naturally, Francesca Vasile, Sara Sattin and Monica Civera were always willing to help and good reference points in the workings of UniMi. There is a very special thank you for the students that have put their efforts next to mine for this project: Nicolò, Davide and Daniele. I hope I've helped you as much as you have helped me: it's been very gratifying to see you become increasingly skilled scientists. I cannot thank you enough: the successes in this thesis also belong to you. Finally, in Milan I had the luck to work closely with my great collaborator, Kanhaya, and also Nives, of the PhD4GlycoDrug family.

Coming back to Grenoble, I was happy to see familiar, but also new faces, such as Simona, who I shared the office with, Sue, Dylan and eventually Jalaa and Nathan. All of them have helped me in one way or the other to press on forward with this project, many thanks to you guys! Also in Grenoble, I was able to work alongside Kanhaya and Margherita of the PhD4GlycoDrug family.

Indeed, the PhD4GlycoDrug is a true family in which the ESR's are true relatives (I stopped short from writing siblings!). Elena, Bene, Tiago, Sjors, Gabri, Mujtaba, Cyril, Marghe, Nives, Kanhaya and Dania: what a bunch of amazing people, I wish I could write five paragraphs to each of you! Thank you so much for being my second family in Europe, I truly wish that we never get too far-off each other! Kanhaya: thank you so much for the effort and time you have spent working with me, and the good times spent all around Europe. Josh will never be higher! Nives: thank you for showing me that a non-italian can also become a perfect Milanese, I swear I'll stop wasting capillaries one of these days! Margherita: BRRRRR! Really: always good times, what else to say? (on celery mode!). And Dania: <u>The</u> colleague.

Dania has been almost as much of a teacher as a colleague. We started this together; and we'll finish it together. She knows as much as I need to learn, and her point of view is always the good complement to mine. I hope I have helped you as much as you have helped me, but I doubt it, because it's so much. I don't believe I'll be able to find another person like you, with whom to fight about everything (and laugh about everything). I wish you the best of lucks, and to never stop growing! Rodrigo is ESR13, no discussion!

I will have forgotten some colleagues along the way, I'm sorry for it, but I'll say this: I haven't had one unpleasant co-worker in over three years. So, you know!

Other people that haven't necessarily contributed scientifically but deserve mention are the old and new friends. From the humble beginnings with Tim and Quentin, passing through Julian, Tom, Stefano, Sergio and Rafa, and getting to Antoine, Aurelien, Damien, and the essentials Rafael, Bob, Guilain, Fepo and Juampi. I wish I would've been less busy these three years. Let's meet soon! There is also a long list of friends to which I owe apologies instead of acknowledgements, hopefully we'll finally meet for a beer in the near future.

Nearing the end, my family has undoubtedly supported me through this period. Mamá y Papá, les debo todo, lo que he logrado es gracias a ustedes. A Cecilia y Rocío, espero abrazarles pronto y gracias por ser fans de mis aventuras. A mis hermanos, Nic y Seb, gracias por las mil ayudas aquí y allá, ojalá pasemos más tiempo juntos próximamente, Gabriel y Daniel, ya mismo nos vemos!! Y al resto de mi familia (incluidos primos y primazos), les extraño y agradezco por creer en mí.

Finalmente, c'è Valen, che ho incontrato a Milano, poi a Leiden, poi in Portogallo, poi a Grenoble, e che spero di poter ancora incontrare ovunque io vada. Grazie per avermi tenuto la mano durante quest'avventura. Adesso che finisco faremo lo stesso per la tua!

LIST OF ABBREVIATIONS

AAT	Anti-adhesion therapy	ITC	Isothermal titration calorimetry
αMeFuc	Methyl alpha-L-fucopyranoside	K_D	Dissociation constant
AMR	Anti-microbial resistance	LDA	Lithium diisopropylamide
Ar	Aromatic	LiHMDS	Lithium hexamethyldisilazide
BCC	*Burkholderia cepacia* complex	LPS	Lipopolysaccharide
CF	Cystic fibrosis	Man	Mannose
COSY	Homonuclear correlation spectroscopy	MDR	Multidrug-resistant
CRD	Carbohydrate recognition domain	MS	Mass spectrometry
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition	NCBI GI	National Center for Biotechnology Information GenInfo Identifier
DCM	Dichloromethane - methylenechloride	nHex	n-Hexane
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin	NMR	Nuclear magnetic resonance
DIPEA	N,N-Diisopropylethylamine	OD	Optical density
DMC	2-chloro-1,3-dimethylimidazolinium chloride	PDB	Protein data bank
DMF	N,N-Dimethylformamide	PSA	polar surface area
DMP	Dess-Martin Periodinane	Pyr	Pyridine
DMSO	Dimethyl sulfoxide	R_f	Retention factor
DSC	Differential scanning calorimetry	rt	Room temperature
ESI	Electrospray ionization	RU	Resonance or response units
EtOAc	Ethyl acetate	SAR	Structure-activity relationship

FP	Fluorescence polarization
Fuc	Fucose
Gal	Galactose
GalNAc	N-Acetylgalactosamine
GlcNAc	N-Acetylglucosamine
Glu	Glucose
HAI	Hospital-acquired infection
hAr	Heteroaromatic
HATU	Hexafluorophosphate Azabenzotriazole Tetramethyl Uronium
Hep	Heptose
hLT	Heat-labile toxin
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass spectrometry
HSQC	Heteronuclear single quantum coherence
IBCWG	International Burkholderia cepacia Working Group
IC_{50}	Half maximal inhibitory concentration
IMAC	Immobilized metal affinity chromatography
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
SSL	Staphylococcal superantigen-like proteins
STD-NMR	Saturation transfer difference nuclear magnetic resonance
SV40	simian virus 40
TBAB	Tetrabutylammonium bromide
TEV	Tobacco etch *virus*
TF	Trigger factor
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TLC	Thin layer chromatography
T_m	Melting temperature
TMS	Trimethylsilyl group
TNF	Tumor necrosis factor
Tol	Toluene
WHO	World health organization

TABLE OF CONTENTS 8

1. INTRODUCTION 13

1.1. The problem of antimicrobial resistance (AMR) 13

1.2. Anti-adhesion therapy (AAT), a possible solution 15

1.3. The role of lectins and carbohydrates in infection 18

1.4. Glycans and Glycomimetics as therapeutic agents 23

1.5. Opportunistic lung pathogens: *Burkholderia cenocepacia* and company 30

1.6. Lectins of *B. cenocepacia*: the BC2L family 34

1.7. The superlectin BC2L-C: state of the art 37

2. RESEARCH METHODOLOGY 43

2.1. Production, purification, and structural characterization of BC2L-C-N_{ter} 43

2.2. Biophysical evaluation of lectins and their interactions 51

3. SCOPE OF THE THESIS 61

3.1. Limits to previous studies 61

3.2. The PhD4GlycoDrug Consortium 61

3.3. Thesis Objective 63

4. A NEW CONSTRUCT FOR BC2L-C-N_{TER} 65

4.1. Summary 65

4.2. Article: BC2L-C N-Terminal Lectin Domain Complexed with Histo Blood Group Oligosaccharides Provides New Structural Information 65

4.3. Further information 81

4.4. Outlook 86

5. DESIGN AND SYNTHESIS OF ANTAGONISTS 89

5.1. Summary 89

5.2. Design of monovalent fucoside antagonists 90

5.3. Modular synthesis of C- and N-fucoside glycomimetics 94

5.4. Fragment functionalization strategy 104

5.5. Proofs of concept towards multivalency 109

5.6. Outlook 112

6. EVALUATION OF ANTAGONISTS 115

6.1. Summary 115

6.2. Validation: STD-NMR 116

6.3. Qualitative evaluation: FP, SPR, and DSC 117

6.4. Quantitative evaluation: ITC 129

6.5. Crystallography 132

6.6. Outlook 137

7. CONCLUSIONS AND PERSPECTIVES 139

8. APPENDIX 143

8.1. Scientific Communication: secondment at Glycopedia 143

8.2. Prediction and Validation of a Druggable Site on Virulence Factor of Drug Resistant *Burkholderia cenocepacia* 166

8.3. Experimental section 175

9. REFERENCES 312

LIST OF FIGURES, TABLES AND SCHEMES

Table 1.1. The race between antibiotic development and AMR 14
Figure 1.1. Schematic representation of bacterial adhesion and the anti-adhesion strategy 16
Figure 1.2. Depictions of the endothelial glycocalyx 18
Figure 1.3. Schematic representation of symmetry in lectins 19
Figure 1.4. Strategies used by pathogens for host recognition and adhesion 21
Figure 1.5. Schematic depiction of anti-adhesion therapy at the glycocalyx 23
Figure 1.6. Examples of monovalent glycomimetics 25
Figure 1.7. Schematic depiction of a multivalent compound on a virus-like scaffold 28
Figure 1.8. Binding of monovalent or multivalent glycomimetics to a hexameric lectin 29
Figure 1.9. Electron microscopy images of *P. aeruginosa* and *B. cenocepacia* 31
Figure 1.10. Structural similarity between LecB, BC2L-A and BC2L-C 35
Table 1.2. Affinities measured by ITC for ligands of BC2L-C 38
Figure 1.11. Crystal structure of the N-terminal domain of BC2L-C 39
Figure 1.12. Hexameric arrangement of BC2L-C and cross-linking 41
Figure 2.1. Diagram of molecular cloning 43
Figure 2.2. Agar gel control of colonies SDS-PAGE purification control 45
Figure 2.3. Solubility phase diagram 47
Figure 2.4. Sitting and hanging drop crystallization setups. 48
Figure 2.5. Examples of two Bravais lattices 48
Figure 2.6. Schematic depiction of data collection through X-ray diffraction 50
Figure 2.7. Schematic depiction of ITC 52
Figure 2.8. Schematic depiction of SPR 54
Figure 2.9. Schematic depiction of DSC 57
Figure 2.10. Schematic depiction of STD-NMR 59
Figure 3.1. The PhD4GlycoDrug consortium 62
Table 4.1. Data collection statistics 81
Figure 4.1. Crystal structure of the BC2L-C-Nt/Le^{y} complex 82
Table 4.2. Summary of the interactions observed between BC2L-C-N_{ter} and Lewis y 83
Figure 4.2. Comparison of binding modes: Lewis y and H-type 1 84
Table 4.3. ITC measurements for carbohydrate ligands of BC2L-C-N_{ter} 86
Figure 5.1. Output from SiteMap analysis of BC2L-C-N_{ter}'s CRD 90
Figure 5.2. Binding poses for the top ranked fragments (KL01 - KL12) 91
Figure 5.3. Ligand design strategy and linkages considered 92

Figure 5.4. Examples final molecules docked on BC2L-C-N_{ter}. 93
Scheme 5.1. Modular synthesis towards β-C- and β-N-fucosides. 95
Scheme 5.2. Synthetic route towards the milestone β-C-fucosylacetylenes 96
Figure 5.5. Diastereomeric half-chair conformers of oxocarbenium ions 97
Scheme 5.3 Synthesis of β-C-fucoside final molecules 98
Scheme 5.4. First attempts to produce ligands featuring the propargylic alcohol moiety 99
Table 5.1. Panel of β-C-fucoside final molecules 100
Scheme 5.5. Synthesis of β-N-fucosides 101
Scheme 5.6. Routes evaluated towards intermediate 30 102
Table 5.2. Panel of β-N-fucoside final molecules 103
Table 5.3. Functionalization of fragments 106
Scheme 5.7. Fragment functionalization through FGIs 108
Table 5.4. KL07 and KL08 derivatives 108
Scheme 5.8. *De novo* synthesis of fragments: achievements and perspectives 108
Figure 5.6. Docking of L-galactose in BC2L-C-N_{ter}'s CRD 110
Scheme 5.9 Synthetic route towards L-β-C-galactosylacetylenes 111
Scheme 5.10 Synthesis of (α,β)-substituted C-fucosylacetylenes 112
Table 5.5. Panel of final molecules obtained 113
Figure 6.1. STD NMR experiment 116
Figure 6.2. Principle of FP and setup for competition experiments 118
Figure 6.3. Fluorescence Polarization experiment 119
Figure 6.4. Schematic representation of an SPR competition experiment 120
Figure 6.5. SPR competition experiments 121
Figure 6.6. Proof of concept for SPR competition experiment 122
Figure 6.7. Evaluation of the BC2L-C-Nter SPR chip against oligosaccharide ligands 123
Figure 6.8. Examples of definitive SPR experiments 124
Table 6.1. SPR and ITC affinity measurements for the panel of antagonists 125
Figure 6.9. Additional SPR experiments 127
Figure 6.10. DSC experiments 129
Table 6.1. (bis) SPR and ITC affinity measurements for the panel of antagonists 131
Figure 6.11. Examples of low c-value ITC experiments 131
Table 6.2 Data collection and refinement statistics 133
Figure 6.12. Docking poses of BC2L-C-Nter ligands. 134
Figure 6.13. Electronic density for synthetic ligands of BC2L-C-N_{ter} 135
Figure 7.1. Design, synthesis and evaluation of BC2L-C antagonists 141

1. INTRODUCTION

1.1. The problem of antimicrobial resistance (AMR)

Across the history of the human race and its progress, many barriers have been met and overcome. One of such, and particularly significant, is the fight of humans against pathogenic microorganisms. Very relevant to current times, pathogenic viruses can rise to become global threats, but so can bacteria. Be it the bubonic plague, tuberculosis, cholera, or others, these names still resonate, echoes of times in which the battle against pathogens was a lost one. In such times, a unicellular organism could singlehandedly decimate a percentage of the human population: for example, tuberculosis peaked in the XIXth century and is estimated to, at that point in time, have killed 14% of humanity (all humans that had ever lived to that point), making it the deadliest bacterial infection in history, so far. In 2019, it still managed to infect 10 million and kill 1.4 million people.[1] Adding to this, bacteria have acted in conjunction with pathogenic viruses, for example during the early XXth century, when the influenza pandemic later called 'the Spanish Flu' left millions vulnerable to opportunistic bacterial pneumonia. This pandemic decimated 5% of the world's population at the time.

The aforementioned dark times came to an end in relatively recent times: as the XIXth century gave its way to the XXth, the rapid development and introduction of many vaccines gave a prophylactic means to fight infectious diseases. More importantly, Alexander Flemming's chance encounter with penicillin in 1928 paved the way for the direct fight against bacterial infections with antibiotics. Penicillin's widespread use started in the 1940s during World War 2, and was followed by the 'Golden Age' of antibiotics (1950-70s), humanity's highest point in the fight against microbes. Yet, by 1955, antimicrobial resistance (AMR) to penicillin was a fact only twelve years after the start of its extensive use, as Flemming himself had predicted. Thus, AMR loomed large over modern medicine and scientists, who kept finding new antibiotics, hoping to keep ahead in the race between humans and AMR pathogens (see **Table 1.1**).[2]

Now, at the beginning of the XXIst century, there is no denying it: we are losing the antibiotics race. As seen on **Table 1.1**, the most recently discovered antibiotics (Daptomycin in 2003 and Ceftazidime-avibactam in 2015) lasted only one year before resistance appeared and was documented.[3] Names such as MRSA (Methicillin-resistant *Staphylococcus aureus*) have

reached the general public and names such as 'superbugs' have been coined for MDR and PDR (Multidrug- and Pandrug-resistant) bacteria.

Development		Resistance	
Antibiotic	**Year Released**	**Year Identified**	**Resistant Strain**
Penicillin	1941	1942	Penicillin-resistant *Staphylococcus aureus*
Vancomycin	1958	1988	Plasmid-mediated vancomycin-resistant *Enterococcus faecium*
Amphotericin B	1959	2016	Amphotericin B-resistant *Candida auris*
Methicillin	1960	1960	Methicillin-resistant *Staphylococcus aureus* (MRSA)
Extended-spectrum cephalosporins	1980 (Cefotaxime)	1983	Extended-spectrum beta-lactamase-producing *Escherichia coli*
Azithromycin	1980	2011	Azithromycin-resistant *Neisseria gonorrhoeae*
Imipenem	1985	1996	*Klebsiella pneumoniae* carbapenemase (KPC)-producing *Klebsiella pneumoniae*
Ciprofloxacin	1987	2007	Ciprofloxacin-resistant *Neisseria gonorrhoeae*
Fluconazole	1988	1988	Fluconazole-resistant *Candida*
Caspofungin	2001	2004	Caspofungin-resistant *Candida*
Daptomycin	2003	2004	Daptomycin-resistant methicillin-resistant *Staphylococcus aureus*
Ceftazidime-avibactam	2015	2015	Ceftazidime-avibactam-resistant KPC-producing *Klebsiella pneumoniae*

Table 1.1. The race between antibiotic development and AMR. Adapted from the U.S. Centers for Disease Control and Prevention.[3]

Cases of patients infected with superbugs resistant to 'last-resort' antibiotics such as colistin have already surfaced in 2016.[4] These pathogens, resistant to most of the existing therapies, are especially threatening to hospitalized patients that present risk factors. Risk factors include medical conditions such as cancer, diabetes and immunosuppression, for example, due to chemotherapy. Additionally, immunodeficiency due to either physiological stress (for example, skin damage or malnutrition) or old age can render a patient prey to these pathogens, in what is called an 'opportunistic' infection.[5-7] It is evident that the mere presence of these pathogens in medical environments could quickly turn into a worst-case-scenario: fragile patients threatened by untreatable bacterial infections. Indeed, MDR microorganisms already represent the leading cause of death by hospital-acquired infection (HAI).[7-8]

Undeniably, HAIs by resistant pathogens can grow into a bigger problem, to the point of reversing years of advances in modern medicine. This issue is illustrated by the situation of

patients afflicted with cystic fibrosis (CF). A well-studied genetic disease, CF is caused by a mutation of the cystic fibrosis transmembrane conductance regulator (CFTR). This dysfunction results in thick mucus accumulating in different organs. Its chief consequence is progressive respiratory problems and increased susceptibility to lung inflammation and infections. Although no definitive cure exists, regular advances of modern medicine have enabled specialized treatment and care for CF patients, leading to an overall improvement of their quality of life. In terms of life expectancy, children born with CF in 2021 are expected to live 20 years more than the previous generation of patients.[9] Despite of this, the main cause for morbidity and mortality (at least 80%) in this population are bacterial respiratory infections. Indeed, the thick mucus characteristic of CF translates into a reduced capacity for airway cleansing, making the lungs an ideal breeding ground for opportunistic pathogens.[10] Due to this, CF is considered a high-risk factor in the context of HAIs and is regarded as the main responsible for mortality among genetic diseases in the Caucasian population. Similar to antibiotics, CF patients are losing the battle against MDR pathogens.

Proportional to what is becoming one of the main challenges of the XXIst century, a coordinated response against AMR has been erected at the highest levels: the World Health Organization (WHO), the European Commission and the United States' CDC (Centers for Disease Control and Prevention) all have action plans to implement against the rise of MDR pathogens.[3, 11-12] These plans provide solid advice on how to reduce resistance by better handling of antibiotics, but also highlight the necessity for alternatives in this fight. Therapies involving vaccines, antibodies and bacteriophages are some of the alternatives presented. Another alternative to antibiotics, less conventional but more relevant to this thesis work, is anti-adhesion therapy.

1.2. Anti-adhesion therapy (AAT), a possible solution

In order to act efficiently, infective pathogens need to interact with their environment. First and foremost, a virus or a bacterium needs to recognize the cells of its host in order to start the infective process. At this point it becomes necessary for the pathogen to remain in close vicinity to its host cells. In this vicinity, pathogens thrive: enhanced access to nutrients, shelter from cleansing mechanisms such as airflow or liquid flow, cover from immune factors,

all converge to facilitate infection. Consequently, host-adhesion is a determinant factor in the infective process.

The concept of anti-adhesion therapy surfaced in the 90s and consisted in using monoclonal antibodies as tools to disrupt adhesive interactions between leukocytes and endothelial cells. Disrupting those interactions was therapeutically beneficial in models of inflammation or immune response.[13] As it stands today, AAT still aims to disrupt adhesive interactions, but has broadened its scope considerably. One of its main applications is relevant to our study: to disrupt the interactions between invasive pathogens and their hosts (see **Figure 1.1**).

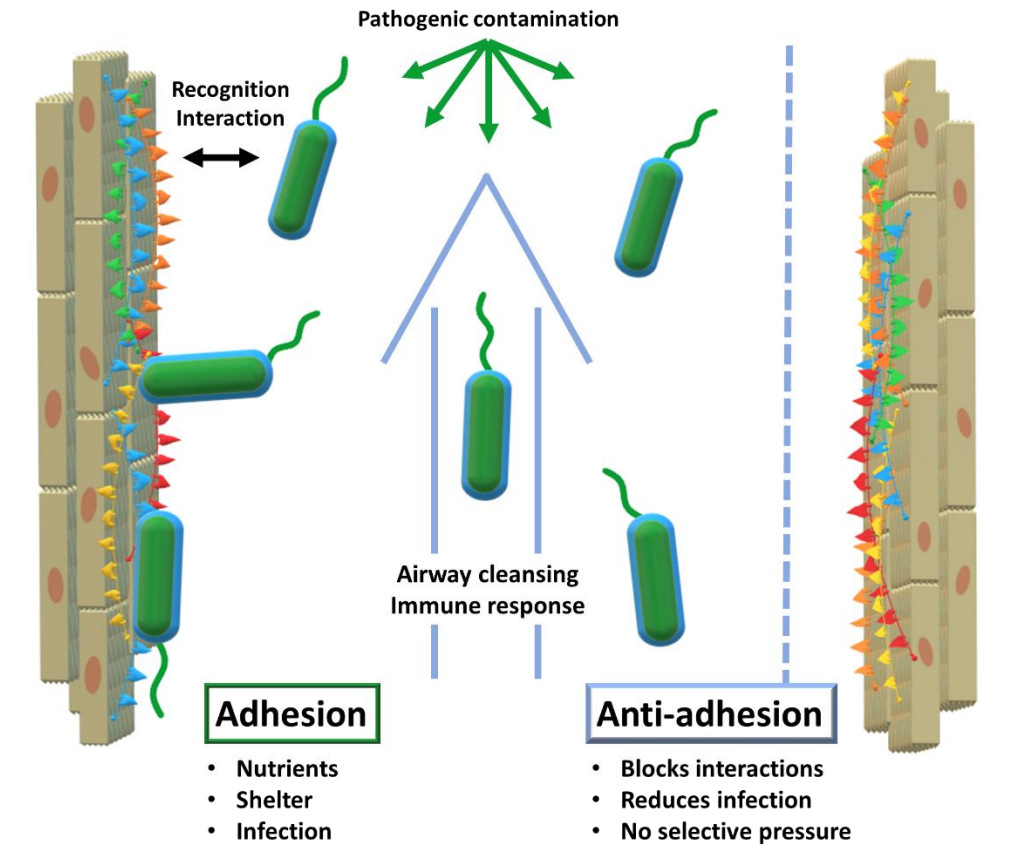


Figure 1.1. Schematic representation of bacterial adhesion to epithelial cells and the anti-adhesion strategy. A detail of the workings of anti-adhesion therapy is presented in **Figure 1.5**.

Indeed, with antibiotic use confronted to increasingly difficult challenges and emerging fungal pathogens developing drug-resistance, anti-adhesion has gained momentum as a complementary type of therapy. The reason why AAT can be deemed complementary to antibiotic therapy is its lack of evolutionary selective pressure: whereas antibiotics only allows drug-resistant mutants to survive and constitute the next generation, AAT doesn't result in elimination of the pathogens. By merely obstructing the infective action of the pathogens, this type of therapy doesn't induce selective pressure in such a direct way. Nevertheless, it

can be argued that mutant organisms that evade AAT and proceed to successful infection will gain an evolutionary advantage: enhanced access to nutrients and capacity to multiply, especially for viruses. On the other hand, these 'favoured' strains will have to compete with normal strains, instead of being the sole survivors of their generation, as opposed to the unhindered growth of antibiotic-resistant strains. The end result is that the resistance to AAT is possible but on a different scale than the dramatic race observed for antibiotics.[14] Furthermore, the prospects of AAT will certainly benefit from the lessons learned from antibiotics, such as the need to limit over-prescription and encourage combination therapies, among others.

Indeed, combination therapies may be instrumental to curb otherwise unsurmountable MDR pathogens. Recently, modern computational tools have been used to model and predict outcomes of combination therapies on simple disease models. Encouraging results showed that antibiotics and anti-adhesives combine synergistically, generating better outcomes than the isolated treatments would. Furthermore, the study allowed to optimise the treatment to arrive to a predicted 'best-case' outcome, in which the minimum antibiotic dose was lower, reducing the chances of resistance to develop.[15] It would, thus, seem that AAT coupled with the gathered knowledge and the newest technologies has the potential to turn the tide in the fight against pathogens.

Among the different AAT approaches against infections, some highlights include the disruption of biosynthesis of adhesion factors of either pathogen or host, the use of antibodies targeting adhesion factors, the immunization of patients against adhesion and the competition against binding epitopes by tailored therapeutic agents.[16] We will develop this last example: the design of AAT agents intended to mimic and compete against epitopes that are usually targeted during the adhesion process in the context of early infection.

As mentioned earlier, adhesion is a staple of infection, meaning that adhesion machinery has evolved throughout time and become increasingly effective and varied. This machinery has also gained specificity in its variety: many different virulence factors specifically target their corresponding epitopes in the host/pathogen interface. Consequently, an understanding of these virulence factors, their targets and the host/pathogen interface is necessary in order to attempt AAT. One key element of this very interface is the so-called glycocalyx: a carbohydrate-populated matrix that encapsulates different types of cells, including epithelial and bacterial cells.

1.3. The role of lectins and carbohydrates in infection

At the forefront of the human anatomy, human epithelial cells separate the body and its cavities from the exterior environment. Their glycocalyx nanolayer is composed by glycoconjugates: glycoproteins and glycolipids which present their carbohydrate portion to the extracellular environment. The role of the glycocalyx and its actors is to sense and communicate with their environment in different ways. For example, epithelial cells are the gatekeepers of the body compartments and, as such, need to communicate to establish a stable cellular tissue. This endothelial tissue assembly is ensured by glycocalyx-mediated communication.[17] Another example of this communication is how glycoconjugates mediate immune self-recognition, allowing the immune system to discern between own and foreign cells, and act accordingly. Finally, the glycocalyx can be a biomarker of diseased states such as cancer.[18-19] Theoretically, the structural versatility of glycans allows them to hold an unfathomably large quantity of information. In reality, this information is filtered through physical and biological constraints, resulting in the glycan structures observed in living organisms. The resulting information held by these glycan structures remains vast: the 'sugar code' is considered the 3rd alphabet of life, employing monosaccharides as letters in parallel to nucleobases and amino acids.[20-22]

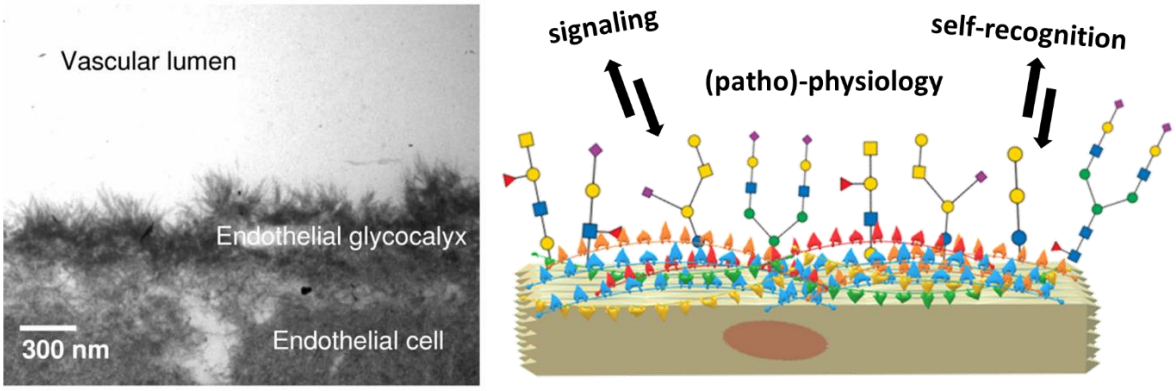


Figure 1.2. Left: Electron microscopy picture of the endothelial glycocalyx. **Right:** Schematic depiction of the glycocalyx and some of its roles. Glycoconjugates and oligosaccharide epitopes are schematized at the surface of an epithelial cell. Adapted from Zausig and co-workers (2013).[23]

Naturally, for every glycan presented by the glycocalyx as a 'message' to its environment, another biomolecule plays the complementary role of 'reader'. Lectins are ubiquitous carbohydrate-binding proteins, key recognition agents for intercellular interactions at the extracellular matrix. Lectins have been studied extensively, owing to their role and potential

for deciphering the sugar code and provide valuable knowledge over its significance on biological processes.[22, 24] Generally having weak millimolar affinity for the monosaccharide version of their ligand, lectins compensate by establishing multivalent interactions, mediated by the presentation of several binding sites. Indeed, lectins often present elements of structural symmetry: β-propellers, β-trefoils and β-sandwiches in homo-multimeric assemblies aren't uncommon. As lectins typically rely on multivalent interactions, they present their binding sites, also called carbohydrate recognition domains (CRD), on the same face. All things considered, the prototypical lectin presents many equivalent or quasi-equivalent CRDs on one of its faces, around a symmetry axis as seen on **Figure 1.3**. Although this seems to imply that lectins have low structural diversity, the opposite is true: lectins hold the structural diversity to match the sugar code. Indeed, the richness of specificity and topology observed in lectin scaffolds have made them interesting tools for generating engineered 'neolectins' with applications in diagnostics, therapy and material science, among others.[25-26] Developed in recent years, UniLectin3D is a valuable database for exploring and comparing lectins and scaffolds: it curates lectins by structural features, but also by carbohydrate specificity and even species, highlighting that lectins are ubiquitous in nature.[27]

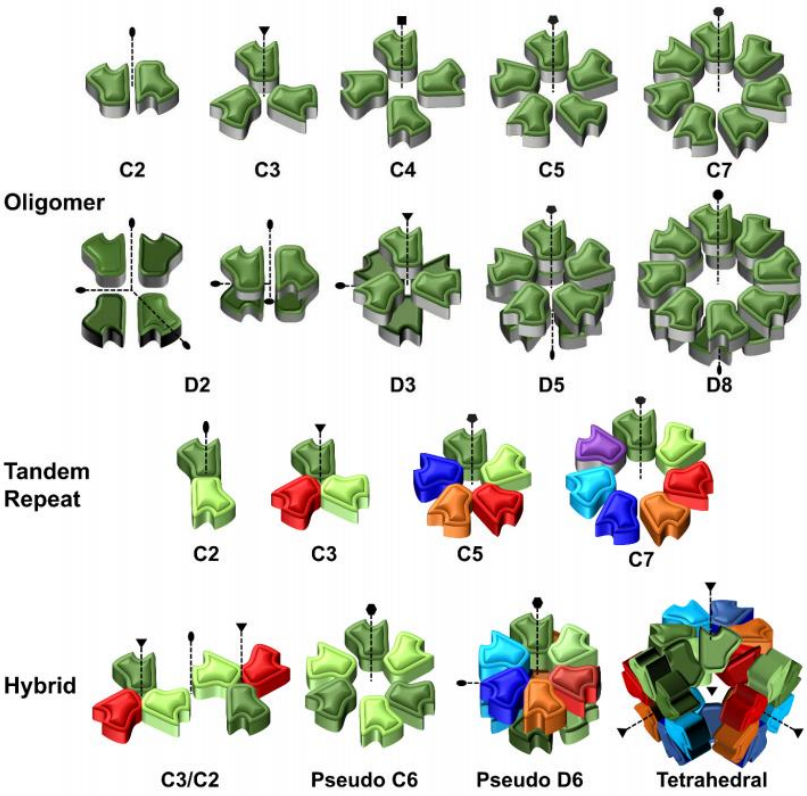


Figure 1.3. Schematic representation of different types of symmetry observed in lectins. The symmetry family and symmetry axis are noted for each schematic representation. Adapted from Notova and co-workers (2020).[25]

Although intercellular communication is not exclusively mediated by lectins, these proteins are particularly represented in the interactions between human and microbes. As mentioned earlier, a pre-requisite to attempt AAT is thorough understanding of microbial virulence factors and their targets in the host/pathogen interface. Belonging to bacteria, viruses, fungi and even parasites, carbohydrate-binding molecules (lectins, toxins, adhesins) are famously known to be virulence factors. On one hand, adhesins are found atop bacterial extracellular organelles – fimbriae, and mediate adhesion of the whole bacterium to any surface that exposes the corresponding carbohydrate epitope. For example, FimH is an extensively studied adhesin which allows *Escherichia coli*'s fimbriae to adhere to mannosylated residues on human epithelial cells, thus facilitating urinary tract infection (UTI). Recently, mechanical studies performed by atom force microscopy (AFM) have been able to characterize the interactions of FimH and other adhesins as 'catch bonds': interactions that get stronger under mechanical tension.[28] The mechanical strength observed supplements another characteristic of these virulent interactions: whereas animal and plant lectins usually have low affinity for their targets, microbial lectins and adhesins present sub-micromolar or stronger affinities.[29]

On the other hand, toxins and lectins are, contrary to adhesins, soluble. Toxins are proteins that usually feature different sub-units. They are released by the pathogen to recognize epitopes on the surface of target cells, which is mediated by a first sub-unit. Upon binding, toxins are internalized, and their second sub-unit enacts a toxic effect, often leading to cell death. A classic example of such toxins is seen in **Figure 1.4**: the AB_5 toxin family. AB_5 toxins featured in organisms such as *E. coli* and *Bordella pertussis* present a cytotoxic ADP-ribosyltransferase (A) domain linked to five (B_5) lectin subunits with capacity to recognize endothelial surfaces.[29-30] Finally, a number soluble lectins don't fill the role of either adhesin or toxin. These agents often present specificity to epitopes presented by the glycocalyx but are not reduced to these targets. Lectins are versatile and can fill complex roles related to quorum sensing, biofilm formation and even cooperativity across different species of pathogens.

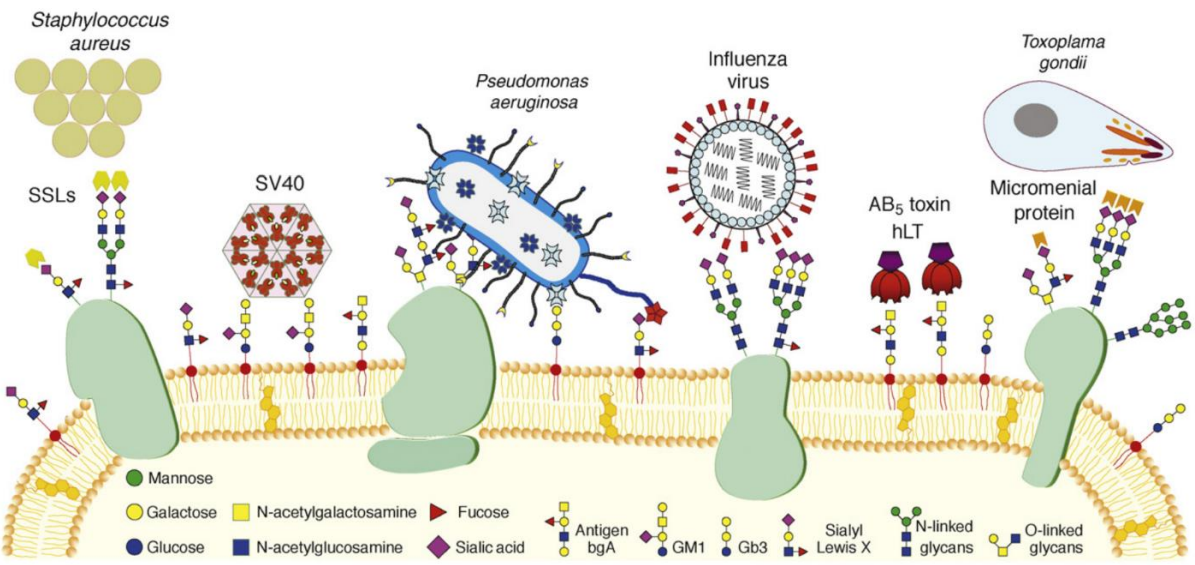


Figure 1.4. Strategies used by pathogens for host recognition and adhesion. Adapted from Imberty and co-workers (2008).[31]

The list of pathogens using lectins for adhesion, infection and toxicity is long: *E. coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Vibrio cholera*, *Clostridium tetani*, *Influenza* viruses, etc.[31] As transpires from **Figure 1.4**, many illnesses and pathologies rely on lectins in their initial stages, cementing the idea that AAT would be beneficial to counter MDR pathogens.[32] What's more, lectins have been shown to have a role in establishing and holding biofilms together, thus boosting resistance to antibiotics. Biofilms are created when bacterial of fungal cells adhere to a surface and to each other to form an extracellular matrix. For pathogenic bacteria, the advantages of forming biofilm are many: stability for growth, change into an infection-adapted phenotype, elasticity against physical forces and, more importantly, resilience against host immune factors and antibiotics.[33] Interestingly for AAT, the knock-out or inhibition of biofilm-mediating lectins has led to disruption of biofilm integrity.[34-36]

Considering this, bacterial lectins are twice-verified targets for AAT: antagonizing all pathogenic lectins would certainly be therapeutically advantageous in the context of early infection. However, every project targeting lectins must be unique: most carbohydrate/lectin interactions are specific. Indeed, lectins are as diverse as carbohydrate structures are. Nevertheless, trends do exist in the context of microbial virulence factors and infections.

Among the common targets for lectins, the role of histo-blood group oligosaccharides in microbial infections is undeniable.[37] Human oligosaccharides are tightly bound to infection, to the point that evolutionary strategies have developed around them. A clear example of this can be drawn from the staple of mammalian biology: breastfeeding. High concentrations

of oligosaccharides are found in the milk of humans and other mammals: they are known as HMOs (human milk oligosaccharides). Interestingly, these HMOs present the same epitopes usually recognized by virulence factors. Studies analysing the influence of HMOs in pathogenicity showed better outcomes for breast-fed infants.[38] This means that by the mere action of breastfeeding, mammals confer a true anti-adhesion therapy to the next generation. Returning to the epithelial glycocalyx, the histo-blood group oligosaccharides present large yet well-defined epitopes for lectins to recognize, which explains the high diversity of microbial lectins and the high specificity for their targets.

As an alternative to microbial lectins, some pathogens take the contrary approach and display carbohydrates that can be recognized by human lectins. By high-jacking human bio-machinery, they are able to infect and, in the case of viruses, enter the human cell in question. Among the pathogens using this strategy is the well-known HIV virus: among others, it targets Langerin and the Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN). This receptor belongs to the immune system and is able to recognize mannosylated glycans characteristic of invasive pathogens such as Ebola, Hepatitis C and HIV. By binding to these viruses, dendritic cells are able to travel to lymph nodes and elicit immune responses. However, HIV takes advantage of this dynamic to propagate and find its way to lymph nodes. Another family of pathogens that has become relevant in recent times uses a similar process: coronaviruses. Indeed, recent studies have confirmed the ability of SARS-CoV-2 to use its spike glycoprotein to target human lectin DC-SIGN and others.[39] We know that the virus enters human cells thanks to the interaction between its spike protein and human angiotensin-converting enzyme (ACE2).[40] It remains to be seen whether the carbohydrate/lectin interactions discovered are also relevant for adhesion and infection.

Returning to AAT, we mentioned the concept of designing therapeutic agents to mimic and compete against epitopes targeted by virulence factors. Applied to virulent lectins, this translates into designing carbohydrate ligands that can compete against the human oligosaccharides by efficiently binding to the lectins, effectively impeding microbial adhesion, as schematized in **Figure 1.5**.

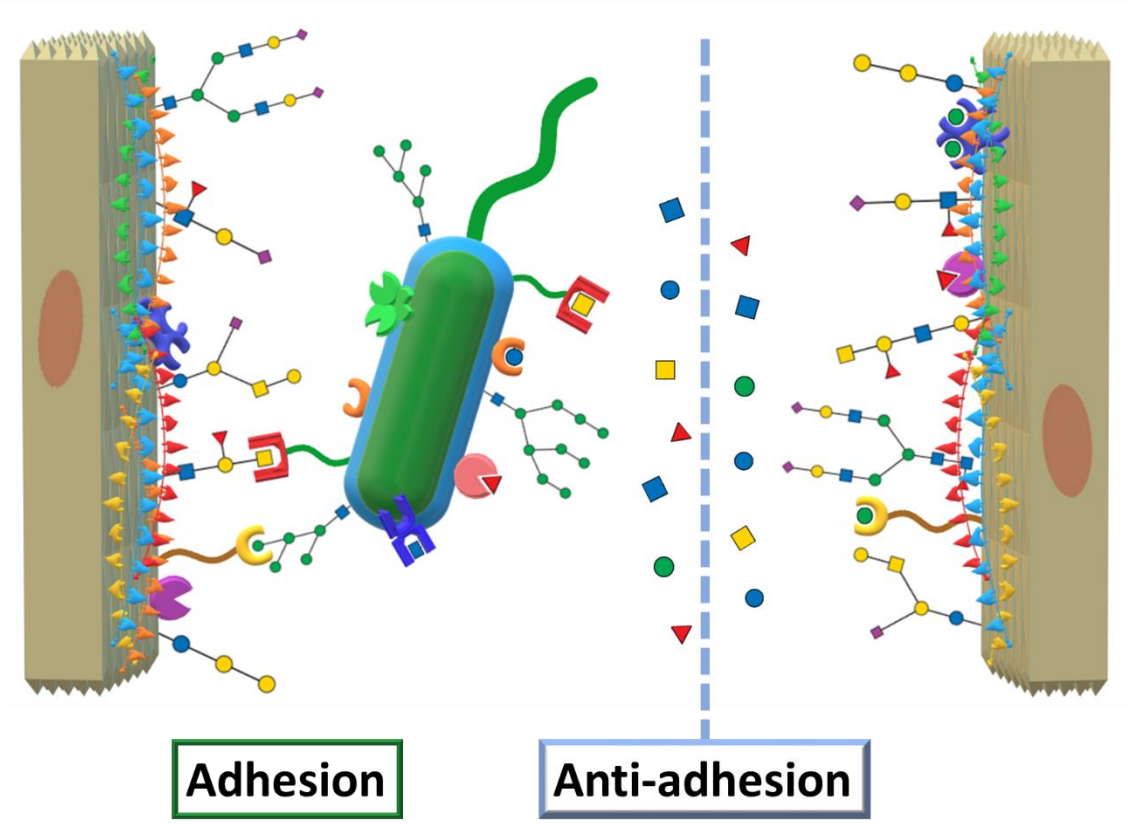


Figure 1.5. Schematic depiction of the carbohydrate/lectin interactions involving epitopes presented by the glycocalyx and the bacterial cell surface and leading lead to adhesion. AAT disrupts the interactions by replacing these epitopes with therapeutic molecules.

1.4. Glycans and Glycomimetics as therapeutic agents

On paper, the concept of using naturally occurring carbohydrates in AAT can be considered a revamping of the successful story involving breastfeeding and HMOs. Unsurprisingly, it was implemented as early as the late 70s and was successful to an extent: *in vitro* and *in vivo* experiments repeatedly prevented infections in models featuring common pathogens such as *E. coli*.[38, 41-42] In animals, successful administration of soluble carbohydrates led to protection of diverse environments: gastrointestinal and urinary tracts, eyes and lungs. Nevertheless, these successes were pushed only to a certain extent, as mono and oligosaccharides showed shortcomings on the prospect of their use for widespread therapy.[38, 43-44]

Indeed, sugars, by their own nature, aren't viable therapeutic molecules. The leading problem is that naturally occurring sugars are 'accounted for' by human biology, meaning that metabolization machinery does a quick job of degrading them into smaller building blocks for recycling. Indeed, sugars can be considered part of the 'building blocks of life', meaning that

a cohort of enzymes exist with the sole purpose of assembling, modifying and dismantling carbohydrate structures. Naturally, this would deplete the effective concentration of any sugar, lowering their therapeutic effect and calling for higher dosage to reach the desired outcome. On a related note, carbohydrate epitopes are found in the glycocalyx and elsewhere, with roles to fill in human biology: communication, immunity or others. This means that overloading the human body with saccharides can have undesired and potentially harmful off-site side-effects.

Another issue with sugars is their large polar surface area (PSA). On one hand, polar molecules are easily dissolved in aqueous solutions, allowing easy administration to epithelial interfaces of infection. On the other hand, large PSA values are usually avoided when designing therapeutic molecules: polar molecules cannot permeate membranes, meaning that some compartments are out of their reach. A good example for this issue are biofilms: if the bacterial targets hide behind a lipophilic matrix, anti-adhesives are as useless as antibiotics. The PSA of a monosaccharide is already high in the scale of drug design: 120 $Å^2$ for glucose, nearing the 140 $Å^2$ upper limit recommended. This clearly means that oligosaccharides are too polar in this scale.

A last argument that separates sugars from drug-like molecules is their low stability: provided they survive enzymatic metabolism, carbohydrates present reactive chemical functions that can easily react with the biological matrix. Furthermore, monosaccharides maintain a dynamic equilibrium between different forms (cyclic forms and open chain). This might not be a problem when it comes to their biological role, but chemical stability is a necessity for a therapeutic entity.

Recapitulating: on one hand, carbohydrates have proven their anti-adhesive potential both in nature and in the laboratory. On the other hand, they fail to align with what modern medicine considers a 'viable' drug. Synthetic organic chemistry has provided a solution to this predicament: glycomimetics. As their name indicate, this relatively new class of therapeutic agents aims to mimic carbohydrates in terms of shape and effect. Their parallel objective is to present an optimized pharmacokinetic profile. The resulting therapeutic molecules, or glycodrugs, boast increased metabolic and chemical stability, specificity for their targets and the ability to be adjusted and re-designed by the means of organic synthesis, to continuously adapt them to new needs.

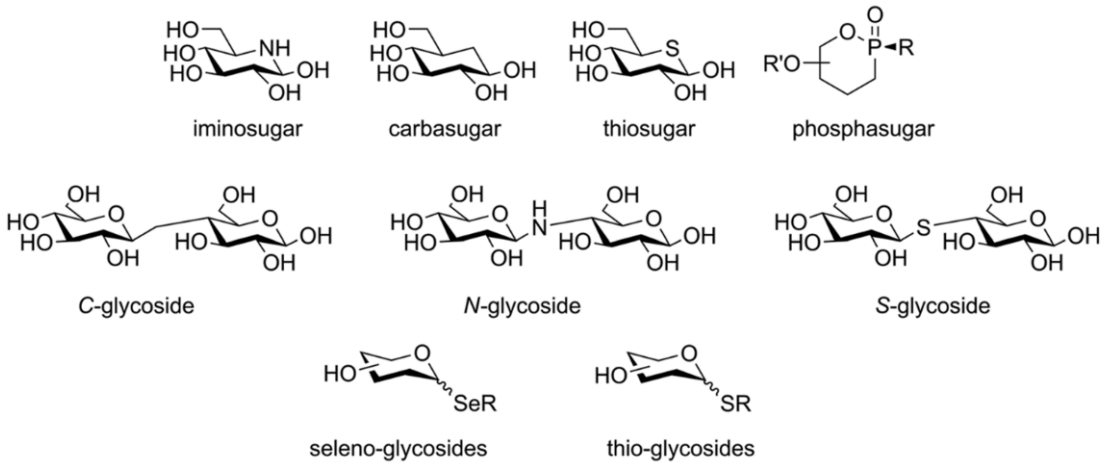


Figure 1.6. Examples of monovalent glycomimetics. Adapted from Tamburrini and co-workers (2020).[45]

Indeed, glycomimetics are molecules tailored to their target: rather than merely copying the original carbohydrate-mediated interaction, they draw on it and try to perfect it. For example, a representative glycomimetic could be modelled after a monosaccharide, but be functionalized with lipophilic moieties that complement the lectin's CRD in order to boost both affinity and specificity for its target. By the virtue of increased lipophilicity, the now moderate polarity of the molecule would grant it access to spaces normally barred for monosaccharides. Finally, the molecule could be synthesized from scratch to replace the ring oxygen by a carbon atom, making it a 'carbasugar' as seen in **Figure 1.6**. This modification would further reduce the PSA, grant it metabolic stability, and also secure the cyclic form from opening. Compared to its monosaccharide equivalent, this hypothetical glycomimetic is already far ahead down the roads of drug-likeness and therapeutic effect.

It's not necessary to go very far to find a real-life glycomimetic success story: carbohydrate-based oseltamivir/tamiflu is a widespread antiviral drug that prevents and treats influenza A and B. Oseltamivir was designed to mimic the transition state generated when the viruses neuramidase cleaves the terminal sialic acid of its substrate glycocompounds. Starting from a slightly modified monosaccharide, synthetic strategies were applied to boost potency and remove structural weak points detrimental to stability or affinity.[46-47] In addition to the installation of a hydrophobic moiety to match an apolar pocket of the binding site, chemical modification also allowed to generate a prodrug derivative, leading to an orally bioavailable glycodrug. Onwards from this early example, the great potential of glycomimetics has sparked a growing number of projects for a range of targets. Some obvious targets for glycomimetics are sugar-metabolizing enzymes: for example, glycodrugs voglibose and miglitol target

glycosidases to achieve glycemic control in the context of diabetes.[46, 48-49] These molecules, an iminosugar and an N-glycoside, are examples of how replacing oxygen atoms by nitrogens can lead to viable glycodrugs. New drugs such as these are always welcome, especially in the case of diabetes: ever-growing pathology of the modern day and the 8th leading cause of death in 2012.[50]

A second significant pathology in which glycomimetics have their role to play is cancer: abnormal cancerous cells exhibit unusual modifications in their glycocalyx, opening an avenue for studying and using cancer-related carbohydrates. Indeed, selectins and galectins are lectin families that have shown involvement with cancer and its aberrant oligosaccharides.[51-52] Many glycomimetic antagonists to selectin and galectin are being developed for cancer combination therapy and are currently undergoing clinical trials.[53-54] Apart from targeting these lectin families, glycomimetics have found their way into cancer therapy in other ways. For example, Gemcitabine is a nucleoside analogue that features a fluorinated ribose mimetic, and has been used in chemotherapy for decades.[55] The list of glycomimetics being developed against these and other pathologies is long, and the synthetic methodologies, ever-growing, as has been recently documented.[45, 56]

Similarly, and closer to our interest, glycomimetics have met success as anti-adhesives. Among many successfully drugged targets, we encounter HIV-related DC-SIGN: based on the oligosaccharide epitopes bound by the lectin, new synthetic glycomimetics have been designed and synthesized throughout the years. They can be separated in the two families recognized by DC-SIGN: mannosides mimicking the epitope Man_9, and fucosides, mimics of Lewis oligosaccharides. Among the many types of glycomimetics designed, high-affinity monovalent structures were created, mirroring the oligosaccharide assembly, yet replacing each sugar by a glycomimetic counterpart.[57-58]

An interesting avenue that synthetic chemistry opens for glycomimetics is that of covalent inhibition: absent in natural structures, reactive groups can be synthetically added to glycomimetics in order to tether them to their targets. This strategy can be applied to carbohydrate-modifying enzymes by taking advantage of their machinery, in what is called *mechanism-based* design.[59] Closer to our interest, anti-adhesive covalent compounds aiming to persistently inhibit lectins have shown promising potential to impede the virulence of the corresponding organism.[60] Nevertheless, the avenue of covalent inhibition entails particular considerations, such as the possibility of unspecific binding and unforeseen side-effects. It

follows that, for covalent glycomimetic design, ensuring selectivity for the target becomes equally or more important than ensuring high affinity. Incidentally, making efforts towards improving the selectivity and affinity of monovalent ligands is a worthwhile step to take before taking glycomimetics to the next level: multivalency.[61-62]

Enabled by synthetic chemistry and its tools, multivalent assembly of glycan ligands has opened the gate to otherwise inaccessible rewards. Ever-increasing numbers of scaffolds and coupling procedures allow straightforward construction of macromolecules bearing repeated units of monovalent ligands. The relevance of multivalent glycocompounds is quite clear: by presenting several copies of the ligand, the monovalent affinities are multiplied to deliver multivalent affinities several orders of magnitude higher. However, this isn't something new: multivalent glycocompounds aim to emulate nature, which usually handles carbohydrate/lectin interactions with multivalency. Indeed, lectins present many equivalent CRDs simultaneously to compensate for low-affinity monovalent interactions. Furthermore, carbohydrates destined for molecular recognition are usually presented in clusters of epitopes, which has been called the 'Cluster Glycoside Effect'. This effect, multivalency, and its implications for therapy have been studied and discussed for decades.[63-66] Some important lessons to retain from the use of multivalent glycocompounds relate to their design and their mechanisms of function.

Regarding design, multivalency has infinite possibilities: glycans have been attached to increasingly large frames, and the valency of these structures has exploded accordingly. Some multivalent designs have completely left behind the idea of drug-likeness in order to produce therapeutic agents at an entirely different scale: carbohydrates supported by nanoparticles, quantum dots, vesicles, micelles, proteins, polymers, and dendrimers have been successfully implemented as tools or therapeutic agents in various projects.[67-73] Pushing design to the limit, virus-like structures bearing over a thousand carbohydrates have been generated, bringing the level of mimicry to a new height (**Figure 1.7**).[74] This infinite potential can, nonetheless, be regulated by some metrics: the geometry of the structure can be defined by the relative orientation of units and the distance between them. Other factors that have a proven influence are rigidity of the construction and, naturally, the number of epitopes presented. Particularly in the case of lectins, it has been established that tailoring the multivalent agent to its target ('lectin-based design') dramatically increases its effectiveness.[63, 66, 75-77]

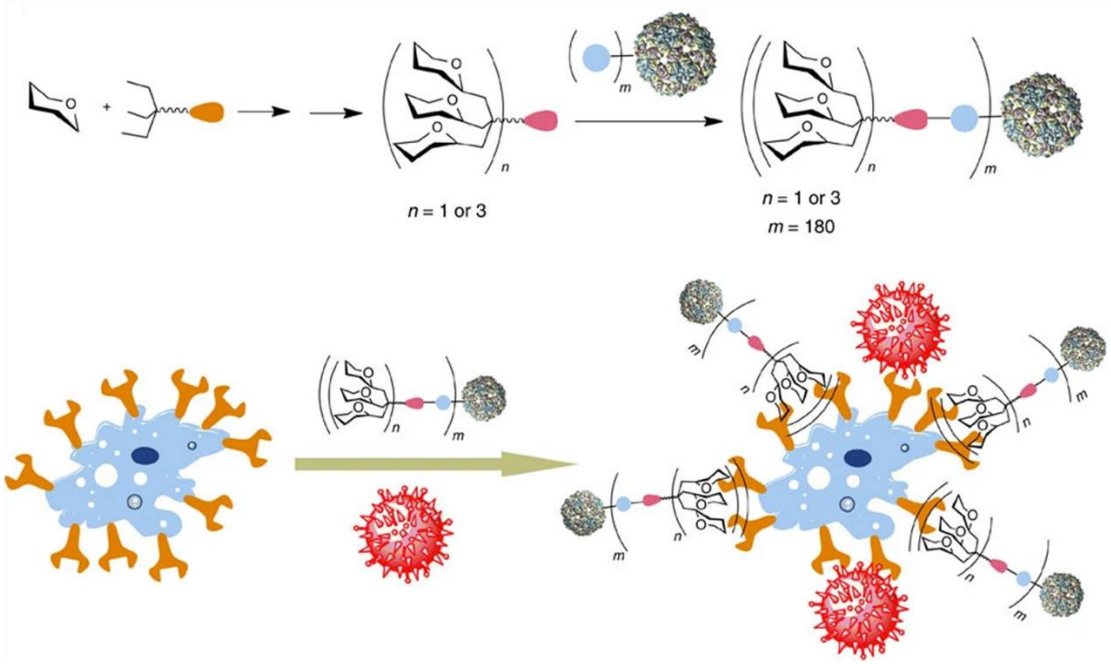


Figure 1.7. Schematic depiction of the nested assembly of a multivalent compound on a virus-like scaffold. The resulting glycodendrinanoparticles are used to compete against Ebola virus in an infection model. Adapted from Ribeiro-Viana and co-workers (2012).[74]

The second lesson to be learned from multivalent glycocompounds relates their mechanisms of function: more than one effect takes place at the same time when these ligands are confronted to their target. Firstly, it is essential to understand what makes an effective multivalent ligand: comparing it to the monovalent unit is useful to assess its affinity and applicability for practical purposes. This 'functional' affinity is called *avidity*, since it is a result of many equivalent interactions, each with their own affinity. It follows that, in order to characterize the efficacy of a multivalent design, it's necessary to correct the multivalent avidity and relate it to a single unit. The comparison of this value to the affinity of a monovalent ligand leads to what could be called a *relative potency* per sugar or per epitope. The increase in relative potency observed when sugars are presented multivalently is the true meaning of the 'multivalent effect'.

With this distinction in mind, it's easier to study the different effects leading to increased affinities and relative potencies, as schematized in **Figure 1.8**. The most intuitive effect is *chelation*, which describes the ability of a molecule to engage two or more binding sites of a target simultaneously. Once a first binding event has established the availability of a multivalent ligand, the affinity of the subsequent interactions is increased compared to the initial binding event. Facilitated binding is one of the drivers of the multivalent effect.

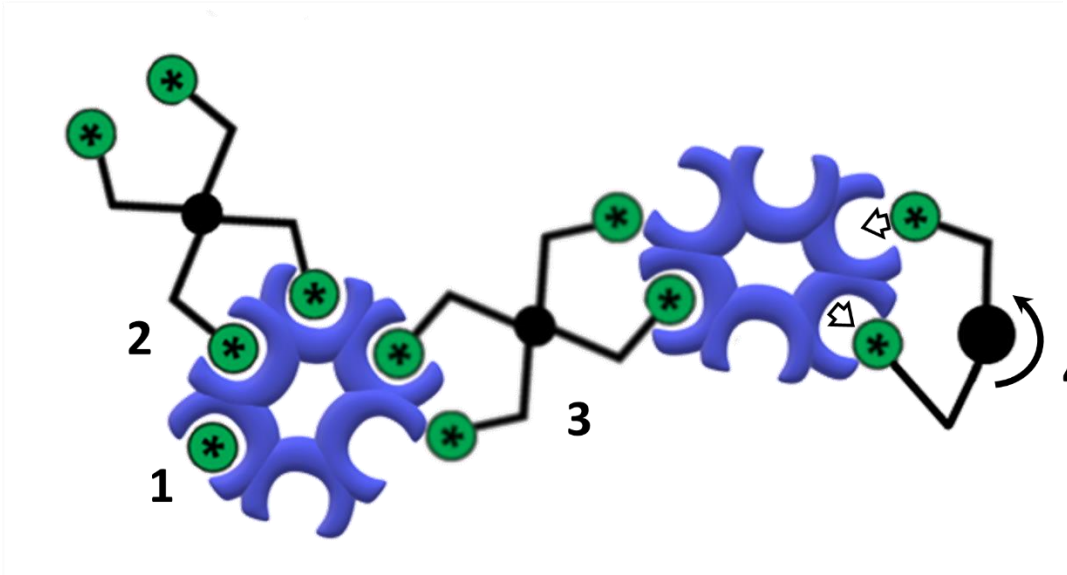


Figure 1.8. Binding of (**1**) monovalent or (**2-4**) multivalent glycomimetic mannosides to a hexameric lectin. Multivalent effects include chelation (**2**), receptor clustering (**3**), and statistical rebinding (**4**).

To push the chelation effect to the limit, multivalent design can envision perfectly tailored compounds that fit to their lectin targets as a lid fits to a pot. Nevertheless, it is a difficult task: any design mistake or fluctuation in the ligand/target dynamics can have dramatic effects on the affinity measured. Parallel to chelation, a second effect called *statistical rebinding* describes the increased likelihood of a second interaction happening on the same site where a first binding event has taken place. This effect works synergistically with chelation and drives the chelation effect even further. Importantly, statistical rebinding can take place also in the absence of chelation: a single site may be consecutively engaged by the multiple copies of the ligand presented in a multivalent structure. Therefore, the off-rate of the ligands is reduced and the affinity, increased. Finally, other effects exist, such as when a compound engages two lectins at the same time, if the steric bulk of the three participants allow it. In this case, the 'recruitment' of targets by a multivalent ligand can be called *receptor clustering* and is known to elicit signalling cascades.[78] In the case of particularly large/long participants with high valencies, cross-linking is possible and can lead to reticulation and even aggregation and precipitation of masses of protein.[65] This aggregation can be beneficial if the aim is to disable the target, such in AAT. It follows that multivalent compounds can also be designed to encompass various targets at once, instead of the 'lid and pot' approach. As of today, the design of multivalent glycomimetics retains a heavy empirical factor, as every target is and behaves differently.

A final word to be said about multivalent glycocompounds is that, although they work well by presenting basic unmodified sugar units, they can benefit from preceding glycomimetic optimization. Indeed, the increased affinity of a monovalent ligand works synergistically: implementing an optimized glycomimetic in a multivalent design can improve its affinity by additional orders of magnitude, as the gain of affinities multiply themselves, rather than adding to each other.[58, 61, 75]

1.5. Opportunistic lung pathogens: *Burkholderia cenocepacia* and company

As previously mentioned, multidrug-resistant (MDR) pathogens are a constant threat to hospitalized patients, especially those with risk factors such as cancer, diabetes, immunodeficiency, etc. Indeed, opportunistic pathogens take advantage of their weakened organisms for infecting and spreading among patients, leading to outbreaks of hospital-acquired infections (HAIs). Among MDR opportunistic pathogens, lung pathogens are especially notorious: lower respiratory infections have been the 4th leading cause of death for the last 20 years.[79] Among the victims of lower respiratory infections, cystic fibrosis patients are particularly vulnerable: in their case, infection by a MDR lung pathogen can easily translate into a death sentence.

Indeed, lung pathogens thrive in the conditions created by CF: thick mucus hinders the action of immune factors and therapeutic agents and reduces the ability for airway cleansing through mucociliary clearance or expectoration. Infected patients often carry infections by one or multiple pathogens throughout years. Even the most invasive interventions such as lung transplantation don't guarantee recovery. What's more: re-infection of the lungs after transplant is not uncommon, meaning that confirmed presence of lung pathogens can be a decision factor leading to denial of this life-saving procedure.[9-10] The list of pathogens associated with chronic lung infection is long: *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Haemophilus influenza*, *Achromobacter xylosoxidans*, *Stenotrophomonas maltophilia*, and members of the genus *Burkholderia*, with more being discovered over time.[10] Among these, two notorious specimens are *P. aeruginosa* and *B. cenocepacia*, albeit for different reasons (**Figure 1.9**).

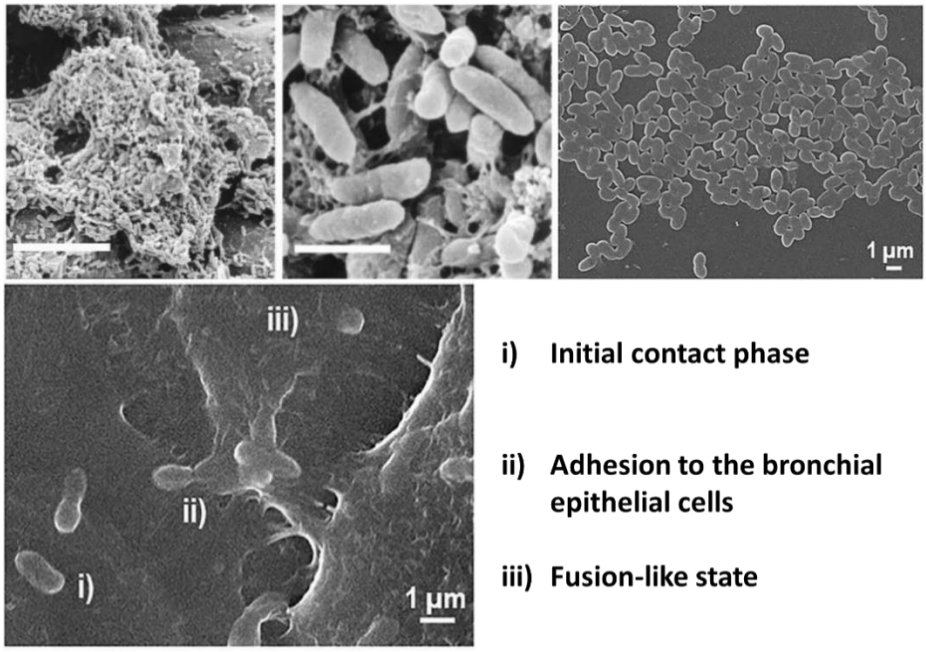


Figure 1.9. Scanning electron microscopy (SEM) images of *P. aeruginosa* PAO1 and *B. cenocepacia* K56-2 cells. **Top left:** *P. aeruginosa* biofilm on a granite pebble, scale: 10µm. **Top middle:** Magnification on *P. aeruginosa* cells, scale: 1µm. **Top right:** Magnification on *B. cenocepacia* cells. **Bottom:** *B. cenocepacia* adhesion to human bronchial epithelial cells (cell line 16HBE14o-). Adapted from Whiteley and co-workers (2001) and Pimenta and co-workers (2021).[80-81]

P. aeruginosa forms part of the ESKAPE pathogens: high-profile threats to human healthcare. In 2017, it was set as a priority for developing new antibiotics by the WHO.[82] *P. aeruginosa* is a Gram-negative bacterium that displays intrinsic drug resistance and easily develops MDR in clinical settings. As such, it is not an inherently easy target to treat, and yet, *P. aeruginosa* lung infections are rarely observed in healthy individuals. This is due to its *opportunistic* behaviour: in humans, infections by this pathogen are seen in conjunction with ailments such as eye injuries, burns, immunodeficiency (AIDS, cancer), and, above all, inflammatory airway diseases (CF, asthma, chronic obstructive pulmonary disease).[83-84] *P. aeruginosa* is responsible for a large part of HAIs (10 % worldwide), and is the main responsible for mortality in CF populations. Two factors are responsible for its high impact: its ubiquitous presence and its capacity to form biofilms. Firstly, this bacterium finds its way to hospitals by virtually every possible path: newly admitted visitors and patients, unsterilized medical equipment, aerators, water supply, and even healthcare personnel.[7] Secondly, once the it establishes itself in a host organism, this pathogen can deploy biofilm and even change its phenotype from 'non-mucoid' to 'mucoid', meaning that it becomes increasingly persistent to antibiotic treatment and deploys factors to boost its virulence.[83-84] This adaptive plasticity is a testimony of how difficult is to permanently eradicate infections by *P. aeruginosa*.

Among its many virulence factors, *P. aeruginosa* produces two widely studied lectins: LecA and LecB, formerly known as PA-IL and PA-IIL.[85] Regulated by quorum sensing, these lectins are released into the extracellular matrix and are known to be essential to biofilm formation, meaning they play a key role in the infection bio-machinery.[34, 36, 86-87] Moreover, both have shown parallel roles in pathogenicity either by mediating cell-adhesion, blocking epithelial ciliary beating or plainly having a cytotoxic effect on lung cells.[88-90]

As such, both lectins have become targets for AAT: *in vitro* and *in vivo* studies demonstrated the usefulness of using the corresponding sugars (galactose, mannose, fucose) for inhibiting the effects of LecA and LecB.[34, 90-91] A small pilot study explored treatment of CF patients by inhalation of monosaccharides: the treatment was well-tolerated and led to promising results, but any claims were limited by the size of the study.[92] Further down the line, mono- and multivalent glycomimetics were developed, with ever-improving affinities and inhibition effects.[93-96] Moreover, as prototypical targets for AAT, these lectins have been used to test innovative strategies such glycomimetic-mediated antibiotic delivery and the first case of covalent lectin inhibition.[60, 97] These recent advancements on the glycomimetic avenue, along with advances in many other fields hold promise in terms of treatment of *P. aeruginosa* infections. Indeed, decades of study and efforts may remove *P. aeruginosa* from its place among the most threatening lung pathogens in the not-so-distant future.

All the information thus presented concerning *P. aeruginosa* can also be related to the main lung pathogen described in this text: *Burkholderia cenocepacia*.

In many aspects, *B. cenocepacia* bears close resemblance to *P. aeruginosa*: it is opportunistic, multidrug-resistant, ubiquitous in the environment, and has led to HAIs in the same way *P. aeruginosa* has.[98] Furthermore, *B. cenocepacia* mediates its infection through quorum sensing, adhesion and virulence vectors.[81, 99-101] Lastly, it has shown the ability to modulate its phenotype during chronic infection and form biofilms, even in cooperation with *P. aeruginosa*.[35, 102-103] On the other hand, some key differences exist between these high-profile pathogens: for instance, *B. cenocepacia* also infects patients suffering from chronic granulomatous disease (CGD).[104] More importantly, its drug-resistance profile is broader and its lung infections are much more likely to spread compared to which of *P. aeruginosa*.[7, 105-106] Adding to this, although *B. cenocepacia* affects less patients than *P. aeruginosa*, its pathogenicity is much more severe and associated with worse patient outcomes. As such, its infections are usually considered more concerning than which of *P. aeruginosa*.[98]

Indeed, *B. cenocepacia*, along with more than 20 strains of the *Burkholderia* genus, have been compiled in what is called the BCC: *Burkholderia cepacia* complex.[107] The BCC was introduced in 1997 by the 'International *Burkholderia cepacia* Working Group' – IBCWG, which was assembled to discuss the emerging threat to public health.[107-108] Among the BCC species, a handful are defined by their role as opportunistic pathogens in lung infection: infection outcomes range from asymptomatic carriage to chronic infection and, in the worst cases, deadly 'cepacia syndrome'. Cepacia syndrome defines a rapid exacerbation of the pulmonary infection: necrotizing pneumonia and septicaemia lead to systemic infection and, if left untreated, death.[105] Although sometimes cured, this syndrome is considered almost untreatable. Due to this, CF patients infected with *B. cenocepacia* and other members of the BCC are often segregated in order to protect other susceptible patients.[109]

In recent times BCC bacteria have sparked severe predicaments: year-spanning outbreaks of *B. stabilis* in Swiss hospitals were studied in 2019, tracing the origin of the contamination to commercially available gloves.[110] Similarly in 2019, the French *ANSM* (National Agency for Medicines and Health Products Safety) had to swiftly release an alert recalling batches of contaminated disinfectant agents.[111-112] This happened after the manufacturer Anios, European market leader in terms of hospital-related disinfection, reported two of their products were contaminated by bacteria: *B. cepacia* and *Pseudomonas oryzihabitans*.[113-114] In this case, the origin of the bacteria was traced to the water supply, highlighting the fact that these ubiquitous bacteria represent an pervasive threat.

Among the species in the BCC, *B. cenocepacia* is multidrug-resistant, it is the species most commonly transmitted among BCC-infected populations, and often accounts for half or more of the total BCC-infections among the studied CF populations.[101, 106, 115] Lastly, *B. cenocepacia* is the main responsible for cepacia syndrome, making it the deadliest species from its genus. This is undoubtedly related to its prevalence, but is also a testament of the particularly aggressive infections that *B. cenocepacia* elicits compared to other members of the BCC.[109] In conjunction, these facts explain why an infection with *B. cenocepacia* is considered most critical and has been studied the most.

Before its reclassification to '*cenocepacia*' in 2003, the pathogenic traits of this species were observed early on as *B. cepacia*'s genomovar III.[107, 116] Extensive study of this bacterium has continued to the present day, from its genome sequencing in 2009 to recent studies dissecting virulence, pathogenicity, existing treatments, and new possible therapies.[81, 99, 101, 117-118] The

current stance in terms of treatment remains antibiotic combination therapy: early aggressive treatment may prevent chronic BCC infections. Nevertheless, in the particular case of *B. cenocepacia,* infections become chronic in over 90% of the cases.[118] Because no consensus on a standardized protocol for exists, the recurring conclusion is that better tools are needed to understand and treat *B. cenocepacia*.

On a different note, pioneering work on gene editing has shown that *B. cenocepacia* and one of its toxins may hold the key to accomplish mitochondrial gene editing.[119] This discovery highlights the importance of exploring a pathogenic target through all the available avenues, which may uncover therapeutic potential or other unexpected applications. As stated previously, recent review articles have explored *B. cenocepacia* and its machinery in terms of determinants for biofilm formation and quorum sensing, adhesins, toxins, etc. Nevertheless, those studies seem to have overlooked the existence of soluble lectins in the proteome of *B. cenocepacia*.

1.6. Lectins of *B. cenocepacia*: the BC2L family

As mentioned earlier, lectins are key actors in cell adhesion leading to infection and have proven to be interesting targets for anti-adhesion and combination therapy. A prime example of these notions is how inhibiting the soluble lectins of *P. aeruginosa* with drug-like glycomimetics has led to biofilm disruption and enhanced susceptibility to antibiotics.[96]

Connecting the dots between *B. cenocepacia* and *P. aeruginosa* is simple: both hold the same characteristics as MDR opportunistic pathogens, target the same populations, are considered critical lung pathogens, and have been extensively studied through the lens of CF-related research. Moreover, both have similar bio-machinery to establish chronic infection: they rely on quorum sensing, adhesion, phenotypic adaptation, biofilm formation and resistance to therapy. Therefore, screening the genome of *B. cenocepacia* and other BBC strains for putative lectins using *P. aeruginosa*'s heavily studied lectins as template can be considered a reasonable venture. The search thus conducted identified four homologs of *lecB* on *B. cenocepacia* strain J2315.[120] The homologs were called BC2L(-A, -B, -C and -D). Three genes in chromosome 2, coding for putative lectins A to C, and a final gene on chromosome 3, coding for putative lectin BC2L-D. Although the gene coding for BC2L-D was invalidated by a frameshift, it was valid in other strains.

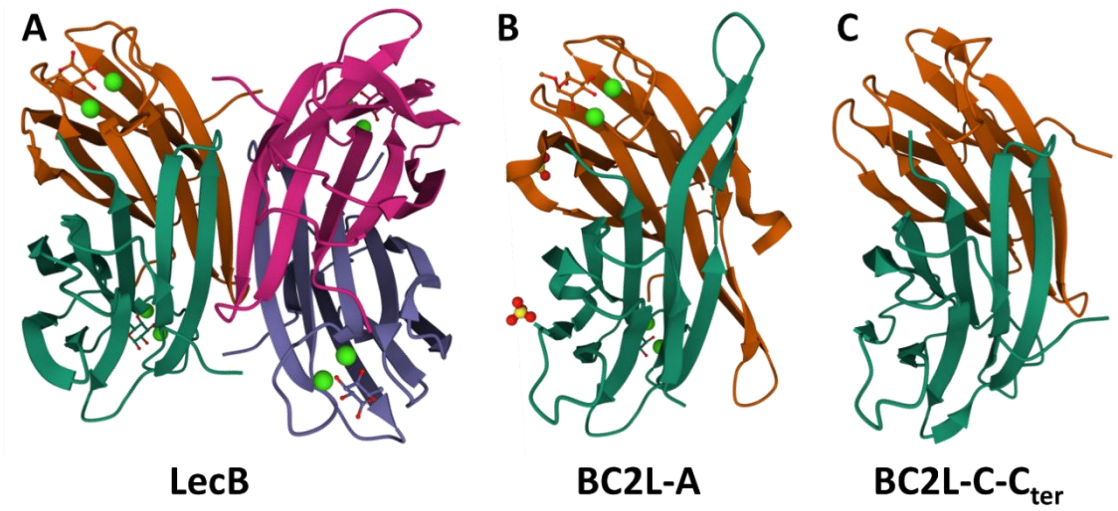


Figure 1.10. Structural similarity between LecB from *P. aeruginosa*, BC2L-A, and the C-terminal domain of BC2L-C, from *B. cenocepacia*. **A:** Homotetramer, ligand is L-fucose (from PDB entry 1GZT). **B:** Homodimer, ligand is αMe-D-Mannoside (from PDB entry 2VNV). **C:** Homodimer, no ligand (from PDB entry 2XR4). Ligands depicted as sticks, ions as spheres: sulfate (SO_4^{2-}, red and yellow), calcium (Ca^{2+}, green).[120-124]

The study of these *lecB*-like lectin family started by BC2L-A, whose original name 'BclA' was aptly changed to avoid redundancy with other protein names such as 'BclA' from *Bacillus anthracis* and the heavily studied 'Bcl-2' family of apoptosis regulators involved in cancer research.[125] Leading to BC2L-A, the *lecB*-like gene *bclA* was found on other six *Burkholderia* strains, well-conserved and maintaining 32% similarity with *lecB*.[120] It coded for 129 residues, longer than LecB mainly through an insertion in a non-functional region and an elongated N-terminus. Nevertheless, BC2L-A was successfully expressed in native form from *B. cenocepacia* strain J2315, and later cloned in *E.* coli and produced in recombinant form, showing the expected LecB-like calcium-mediated specificity for mannoside saccharides.

Indeed, LecB is a fucose-binding lectin also able to bind mannosides and requires two calcium Ca^{2+} ions for carbohydrate binding. Interestingly, BC2L-A shows exclusive specificity for mannosides. This is due to a difference in their sequence: a specificity loop formed by residues 22-24 in LecB features two serine residues (22 & 23), which are replaced by alanine (29 & 30) in BC2L-A, thus allowing rationalization for the specificity. Nevertheless, both lectins have unusually strong affinity for their respective ligands when compared to usual monosaccharide/lectin interactions.[126-127] The structural and functional study of BC2L-A went on to provide crystal structures, extensive probing against mannosides and even successful inhibition with mannoside glycomimetics.[120, 127-130] In these studies, the structural similarities and differences between LecB and BC2L-A were detailed: the homotetrameric form of LecB is inaccessible to BC2L-A, which instead remains a homodimer (see **Figure 1.10**). Additionally,

the potential of glycomimetics as antagonists of BC2L-A was proven through structural and biophysical evaluation, and BC2L-A proved to be a useful model to perform optimization of multivalent glycomimetic design.[131]

A report of outmost relevance described interactions of BC2L-A with epitopes obtained from bacterial lipopolysaccharides (LPS).[132] LPS are structural staples of the Gram negative outer membrane and cover the vast majority of the bacterium's surface.[63] This could mean that the likely biological function of BC2L-A is to mediate cell-cell adhesion between bacterial cells. Finally, fluorescent-tagged BC2L-A was used for imaging experiments: *E.* coli and B*. cenocepacia* cells were incubated with the lectin, which accumulated exclusively at the surface of B*. cenocepacia* and within its biofilm. This study confirmed the ability for this soluble lectin to interact not only with the host mannosylated glycoproteins, but also with bacterial cells, participating in the biofilm matrix.

As the study of BC2L-A advanced, so did the interest in the other orthologs of *lecB*: *bclB*, *bclC* and *bclD*. Indeed, these putative proteins were longer than BC2L-A, featuring N-terminal domains with no relation to LecB. Thanks to their identification as soluble lectins, their role was considered in studies of *B. cenocepacia*'s virulence. One study evaluated the evolution of genomic expression along chronic infection on a single patient who suffered from cepacia syndrome.[102] The transcriptomic analysis surveyed which genes were up- or down-regulated in a period of 3 years of chronic infection. Among others, genes coding for BC2L-B and -C were up-regulated, whereas the corresponding gene for BC2L-A was down-regulated. This difference of outcomes pointed towards the possibility of secondary roles for the N-termini of these lectins.

Other studies detailed the genes regulated by quorum sensing in *B. cenocepacia*.[35, 133] In these, the influence of the lectins -A, -B and -C on biofilm formation was assessed. After proving that the operon *bclACB* coding for the three lectins was regulated by quorum sensing, it was also uncovered that it plays a role in maintaining the structure of biofilm. Indeed, gene knock-out strategies confirmed that biofilm was slower to grow when the lectins were absent and was structurally flawed when compared to wild-type biofilm: it presented cavities and alterations in thickness and biomass. More importantly, lectin-specific knockouts revealed that all three lectins were necessary and the lack of any one of them led to defective biofilm. This discovery hints at specific roles of each lectin, again pointing to the role of the N-termini of BC2L-B and -C. Compared to *P. aeruginosa*, blocking lectin action on *B. cenocepacia*

produced a mitigated effect (malfunction instead of disruption). Nevertheless, this information remains encouraging if these lectins are to become targets for AAT.

Taking into account the growing body of data, it was clear that the discovery of LecB-like proteins featuring additional N-terminal domains was not trivial. The study of these domains could uncover information related to the role of the BC2L family in virulence and the adhesion mechanisms of *B. cenocepacia*. It was so that the lectin BC2L-C and its N-terminal came under close scrutiny.[134] Their study would reveal a *superlectin*.[124]

1.7. The superlectin BC2L-C: state of the art

Similar to BC2L-A, the protein BC2L-C is a "Lec-B" like lectin: its C-terminal domain is 116 residues long and shares 43% identity with LecB.[124] Much like BC2L-A, this domain assembles itself as a homodimeric lectin, featuring two calcium-dependant carbohydrate recognition domains. Continuing the similarities, the specificity loop in the CRDs bears alanine residues, ensuring specific affinity for mannosides and mannosylated structures in the low micromolar range. Nevertheless, beyond the 116 residues of its C-terminal domain, BC2L-C departs from BC2L-A and becomes a unique lectin.

BC2L-C was originally identified from *B. cenocepacia* strain J2315: its gene *bclC* (NCBI-GI 206562055) codes for 272 amino acids, and has been consistently found in other *B. cenocepacia* strains.[120, 134] The C-terminal Lec-B like lectin that led to its discovery accounts for 116 residues. The next 26 amino acids form a serine- and glycine-rich flexible region, which is considered a linker. The remaining amino acids of BC2L-C form its 130 residues-long N-terminal domain. As previously mentioned, lectins presenting many domains are not uncommon: multivalency can be achieved by repeating the same lectin unit. Nevertheless, the N-terminal domain of BC2L-C (BC2L-C-N_{ter}) was found to be a lectin domain structurally different to BC2L-C-C_{ter}, with well-defined carbohydrate specificity for fucosides.[134] This dual carbohydrate specificity is exceptional and defines the chimeric BC2L-C as a *superlectin*.[124]

Two seminal studies characterized BC2L-C. In 2010, Šulák and co-workers expressed the 28 kDa native protein, then designed a gene coding for the 156 N-terminal residues of the protein and cloned it in *E.* coli for recombinant production.[134] Indeed, this initial design included the linker region, which was only characterized as such in retrospective. The protein construct thus obtained was labelled BC2L-C-nt and was 187 residues-long, due to the un-cleavable 31

residues-long C-terminal histidine tag (HisTag), that was engineered for purification. Size exclusion chromatography revealed the first structural feature of this domain: its elution size corresponded to a 58 kDa protein rather than the expected 19 kDa, signifying a homotrimeric assembly. Assuming it to be a lectin, Šulák and co-workers went on to characterize the new construct by probing it against different monosaccharides. BC2L-C-nt showed specific millimolar affinity towards L-fucose by surface plasmon resonance (SPR). To further define specificity, the construct was probed against a glycan array, resulting in marked preference for fucosylated histo-blood group epitopes. Indeed, the N-terminal of the BC2L-C *superlectin* is specific for a well-known lectin target for adhesion, present in the glycocalyx of human epithelial cells. Isothermal titration calorimetry (ITC) allowed further characterization of affinity against human oligosaccharides, returning micromolar values detailed in **Table 1.2**. The best affinity for BC2L-C-nt was found for Lewis y (Le^y: αFuc1-2βGal1-4[αFuc1-3]βGlcNAc), with a K_D of 54 µM. The data collected confirmed one CRD per monomer, meaning three binding sites per trimer.

Ligand	Terminal Epitope	Affinity (µM)	Reference
BC2L-C-C_{ter}			
D-Mannose	Man	37.4	Šulák and co-workers (2011)
αMeMan	Man	27.6	
Trimannose	Manα1-3(Manα1-6)Man	28.8	
αMeHept	L,D-manHep	236	
Diheptose	L,D-manHepα1-3L,D-manHep	88.1	
BC2L-C-N_{ter}			
αMe-L-Fuc	Fuc	2700	Šulák and co-workers (2010)
H-type 2	Fucα1-2Galβ1-4GlcNAc	1236	
Lewis b	Fucα1-2Galβ1-3[αFuc1-4]GlcNAc	213	
Lewis x	Galβ1-4[αFuc1-3]GlcNAc	196	
Lewis a	Galβ1-3[αFuc1-4]GlcNAc	132.1	
H-type 1	Fucα1-2Galβ1-3GlcNAc	77.2	
Lewis y	Fucα1-2Galβ1-4[αFuc1-3]GlcNAc	53.9	
BC2L-C			
D-Mannose	Man	21.8	Šulák and co-workers (2011)
αMeMan	Man	18.3	
Lewis y	Fucα1-2Galβ1-4[αFuc1-3]GlcNAc	47.5	

Table 1.2. Affinities measured by ITC for different ligands of the two domains of BC2L-C. Standard deviations are below 5%. Adapted from Šulák and co-workers (2010 and 2011).[124, 134]

In the same study, the first crystal structure of this domain (PDB-ID: 2WQ4) was solved at 1.42 Å, as seen in **Figure 1.11**. It revealed a trimeric structure presenting β sheets in a jellyroll -

Greek key architecture, which was unprecedented for lectins. It closely resembled the structure of the human tumor necrosis factor (TNF), heavily studied for its role in signalling and immunity, despite no sequence identity. It is worth to mention at this point that the sequence coding for BC2L-C-N_{ter} did not match any other known sequence, with the exception of a putative protein from the unrelated *Photorhabdus luminescens*, an insect pathogen. Three fucoside-populated CRDs were found at the protomeric interfaces, presented on the same face of the structure, which facilitates interactions with surface-bound epitopes, as observed in many lectins.

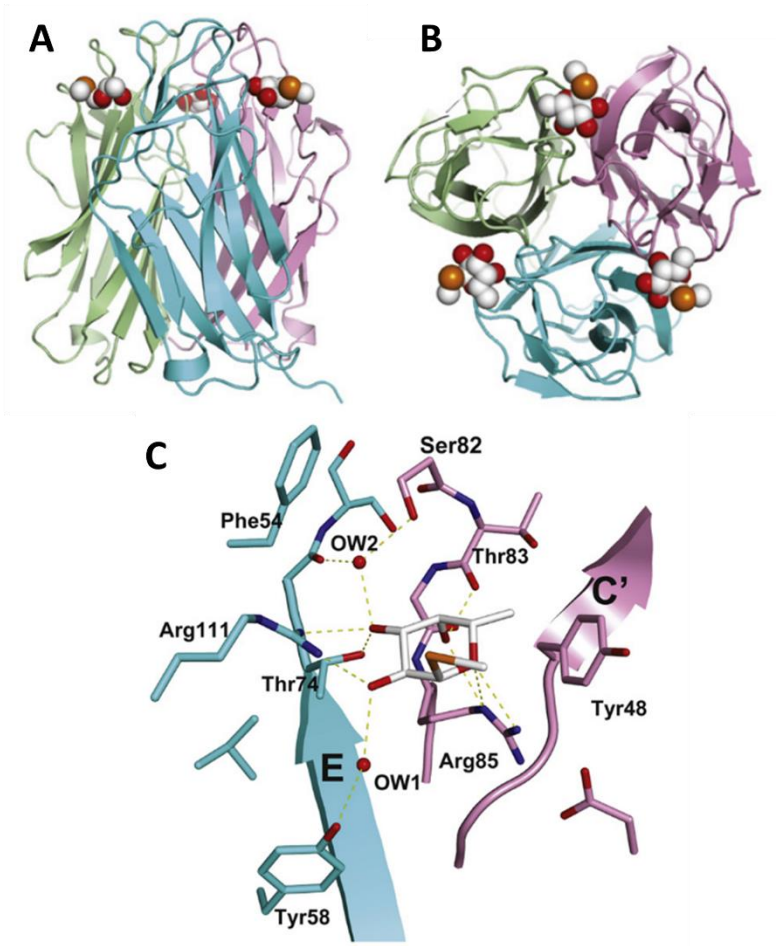


Figure 1.11. Crystal structure of the N-terminal domain of BC2L-C. **A and B:** Side view of the homotrimer, ligand is αMe-Seleno-L-Fucoside. **C:** Details of the CRD of BC2L-C-N_{ter} and interactions with the ligand. Water molecules depicted as red spheres, ligand as spheres or sticks, H-bonds as yellow dashed lines. Adapted from Šulák and co-workers (2010).[134]

Study of the binding interaction allowed to rationalize the observed L-fucose selectivity. Arginine residues Arg111 and Arg85 belonging to the two adjacent protomers, encase the monosaccharide from the side and below, respectively. They establish hydrogen-bonds (H-bonds) with oxygen atoms O2, O3, and O4, O6, respectively. Other noteworthy interactions involve a water molecule buried between ligand and protein, which establishes H-bonds with

the ligand's O3 atom and residues Tyr75 (carbonyl) and Ser82 (side chain). The remaining interactions are detailed in **Figure 1.11**, overall constituting a novel fucose binding mode previously unseen in other lectins. The selectivity for L-fucosides and related L-galacto-configured structures can be condensed to the substituent at the C4 position: residue Arg85 allows the space for a downward axial substituent, whereas an equatorial substituent would generate steric conflict with the side chain of Ser82.

Although the crystal structure was solved at high resolution, the residues corresponding to the linker region weren't visible as a result of disorder and with high mobility. This introduces the main limit of this study of BC2L-C-N_{ter}: the unaccounted flexible tail of the construct was detrimental for stability, with precipitation being a common problem.[135] Similarly, attempts to co-crystallize the protein with larger ligands were unsuccessful, since accessibility to the binding side was probably hindered by the flexible region. Thus, structural information of the carbohydrate/ligand interaction with human oligosaccharides couldn't be obtained, although their affinity for the lectin is several orders of magnitude stronger than the monosaccharide's. Molecular modelling led to predicted binding modes of oligosaccharides such as Le^y and H-type 1, to be verified by future studies.

In 2011, a subsequent study from Šulák and co-workers uncovered the superlectin as a whole: recombinant versions of the C-terminal domain and the full protein were produced and probed. From this, the aforementioned similarities with BC2L-A were defined, and the affinity for different mannoside and manno-configured heptose ligands, quantified (see **Table 1.2**).[124] A crystal structure of the recombinant BC2L-C-C_{ter} dimeric lectin domain was obtained (PDB: 2XR4, **Figure 1.10**), and was used for successful computational docking of mannosides as ligands.[136] Apart from the characterization of the LecB-like domain, the full protein was characterized in terms of affinity and structure. Glycan array, SPR and ITC technologies were reprised to confirm the dual specificity, revealing no overlaps between the binding abilities of both domains. Structural analysis by small angle X-ray scattering (SAXS) and electron microscopy (EM) revealed a flexible hexameric structure, which accounted for the (three) dimeric and (two) trimeric C- and N-terminal domains (see **Figure 1.12**).

This arrangement supported the working theory that the superlectin acts as a 'cellular bridge' or cross-linker of human and bacterial cells by simultaneously engaging with the surface-bound epitopes recognized specifically by each terminal. Indeed, the hypothesis is further supported by three additional findings. Firstly, BC2L-C, along with -A and -B, are secreted by

the bacterial cell into the extracellular medium by a yet unknown mechanism. Secondly, these lectins can be found at the bacterial surface, later confirmed for BC2L-A and -B in separate studies.[35, 132] Lastly, the BC2L-C was released into the extracellular matrix only upon incubation of the cells with mannose, hinting heavily at its regulation by quorum sensing and involvement in virulence.[124]

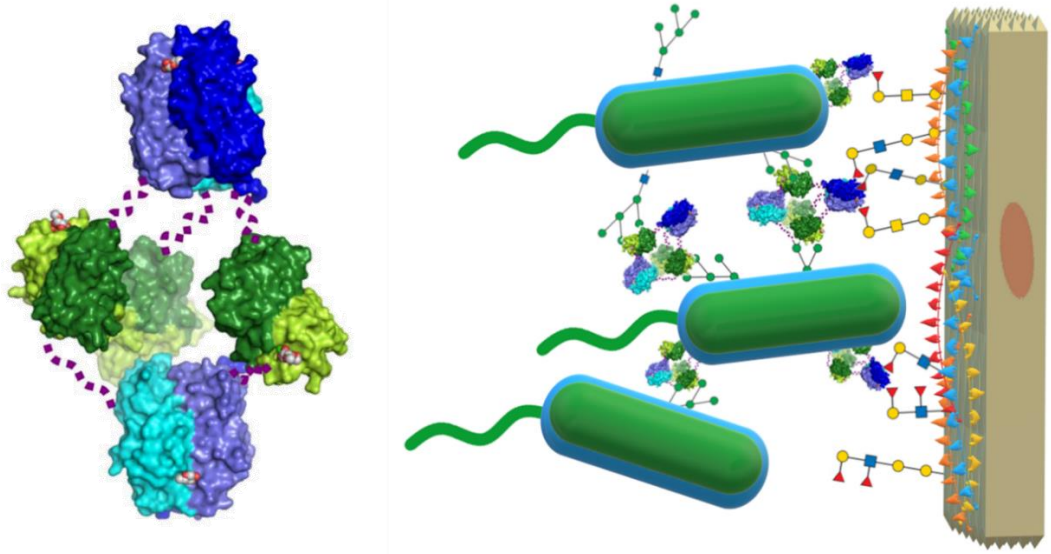


Figure 1.12. Left: Likely hexameric arrangement of BC2L-C from the SAXS and EM reconstructions. **Right:** Schematic depiction of the 'cellular bridge' hypothesis: BC2L-C cross-links *B. cenocepacia* and human epithelial cells by binding their LPS mannoside and histo-blood fucoside epitopes, respectively. Adapted from Šulák and co-workers (2011).[124]

On a different note, this study evaluated the recently discovered structural relation of BC2L-C-N_{ter} to inflammatory elicitor TNF. In particular, the study assessed whether exposing epithelial cells to the superlectin would trigger an immune response. A marked increase in secretion of interleukin 8 (IL-8) upon treatment with the full protein or its N-terminus proved the hypothesis. Nevertheless, the inflammatory pathway remained obscured by the fact that the obvious candidate, TNF receptor 1 (TNFR1), was not engaged by BC2L-C-N_{ter}. Alternatively, it was demonstrated that the inflammatory response wasn't linked to carbohydrate binding. The capacity of a virulence factor from *B. cenocepacia* to elicit inflammation through a cytokine-like structure can be tied to the heavy inflammation seen in patients with cepacia syndrome. Hard proof of the relationship between BC2L-C, inflammation and cepacia syndrome remains to be obtained by further study.

In the decade since its initial characterization as superlectin, BC2L-C has been studied and implemented by many. Across many works Tateno, Ito and co-workers have established the affinity of the lectin for fucosylated oligosaccharide epitopes on human pluripotent stem cells.

Their new construct of BC2L-C-N_{ter}, called rBC2LC-N, was produced and implemented for fluorescence-based techniques to detect induced pluripotent stem cells and embryonic stem cells against differentiated stem cells (iPSCs, ESCs and SCs, respectively).[137-139] The glycoprotein Podocalyxin was identified as a cell-surface ligand of rBC2LC-N, through its H-type 3 epitopes.[140] This discovery allowed the development of a method for the detection and elimination of tumorigenic pluripotent cells, with a direct use for safety in stem cell therapy.[141-142] In recent years, their efforts have led to develop chimeric proteins featuring rBC2LC-N and various toxins: lectin-drug conjugates (LDC), aimed at varied cell targets specifically recognized by the lectin domain.[142-143] Finally, they have highlighted the usefulness of BC2L-C-N_{ter} to detect specific populations of cancer cells.[144-145] Among the many discoveries from this line of research, more support for the 'cellular bridge' hypothesis can be found: BC2L-C-N_{ter} probes bind to human cells via the histo-blood groups in their glycocalyx with antibody-level sensitivity and they seem to specifically bind to cell lines with epithelial characteristics.[138, 143-144, 146]

Apart from the work of this group on stem and cancer cells, others have benefitted from BC2L-C as a tool for varied endeavours: detection of histo-blood epitopes for cell characterization, development of protein stability screening kits, validation of microbe-oriented glycan arrays, and validation of Fragment Molecular Orbital (FMO) tools for the analysis of protein/ligand interactions.[135, 147-150] Finally, some groups have taken the challenge of antagonizing the superlectin with an early array of fucoside ligands, reaching some degree of success thanks to multivalency.[151-152] From these campaigns, the best synthetic ligand for BC2L-C-N_{ter} was a calix[4]arene-based tetravalent fucoside which showed a 256-fold increase of potency compared to L-fucose for inhibition of hemagglutination of red blood cells. A cross-linking test confirmed the capacity of this inhibitor to aggregate *B. cenocepacia* cells by engaging the surface-bound lectin. Bearing mostly unmodified C-fucosides, this compound proves successful multivalent effect: the potency per sugar corresponds to a 64-fold increase. With multivalency validated as a viable strategy to inhibit this virulence factor, S-fucoside glycomimetics were put forward in an attempt to develop glycomimetic monovalent inhibitors. Nevertheless, the optimization of these was hindered by lack of biophysical techniques to assess affinity constants and structural data to rationalize the relative potency observed in hemagglutination assays.[152]

2. RESEARCH METHODOLOGY

The scientific methods used throughout this work can be separated in three kinds:

- Production, purification, and structural characterization of BC2L-C-N_{ter}
- Biophysical evaluation of lectins and their interactions
- Organic synthesis and characterization of small molecules

The principles related to the first two will be described in this section. The protocols and materials employed for experiments will be detailed in the **APPENDIX 8.3: Experimental Section**, along with the characterization of the synthetic molecules.

2.1. Production, purification, and structural characterization of BC2L-C-N_{ter}

Primer Design, PCR and ligation

The experiments described in this work require large-scale production of a lectin. To achieve this, it is necessary to express a recombinant version of the protein in a host organism, typically *E. coli*. However, before expression in *E. coli*, the genetic material coding for the protein must make its way to the host organism.

The first step towards this is to obtain the genetic material coding for the target protein, which will serve as template for the amplification of a new version of the gene (see **Figure 2.1**).

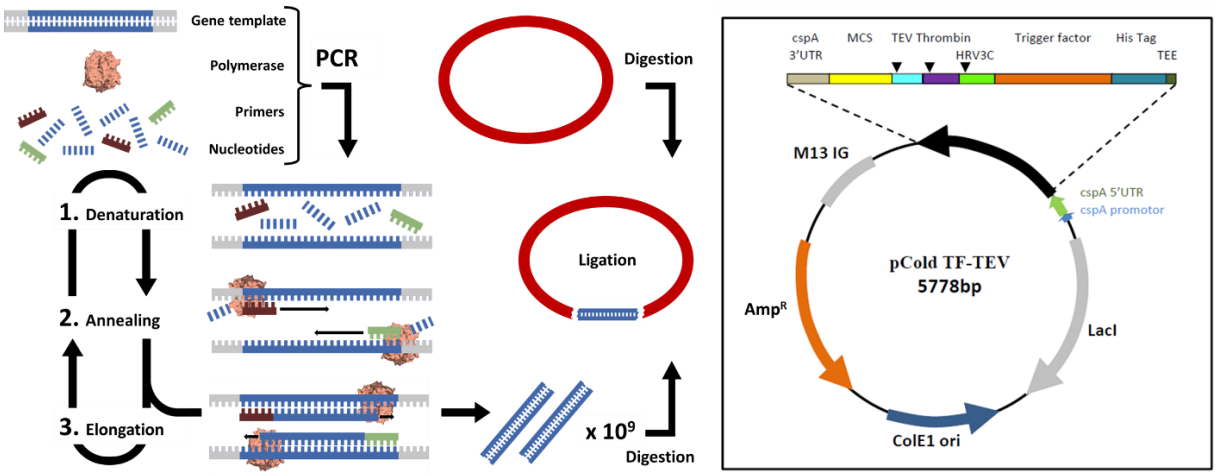


Figure 2.1. Left: Diagram of molecular cloning, from gene template to the recombinant plasmid. **Right:** Plasmid chosen for cloning and expression: pCold TF – TEV.

The second step is to design primers: short single-strand DNA sequences that complement and delineate the genetic material to be amplified. With these two elements, a DNA polymerase and nucleotides are added to start Polymerase Chain Reaction (PCR). PCR consists in the amplification of the gene template, defined by the primers and mediated by the polymerase (see **Figure 2.1**). Through cycles of denaturation, annealing and elongation, billions of copies of the gene are created. Subsequently, the amplification of the gene at the correct size can be controlled by agarose gel. Pressing forward, digestion of the new gene is performed by restriction enzymes: the new restriction sites added by the specifically designed primers are cleaved into 'sticky ends'. In parallel, a plasmid is chosen as cloning and expression vector for the new gene: its multiple cloning site (MCS) is digested for the same restriction sites as the gene. Here, we used pCold TF – TEV (see **Figure 2.1**), modified from pCold TF, Takara Bio Europe.[153]

Ligation of the digested genetic materials (insert and vector) is achieved by mixing them to a ligase, the new recombinant plasmid is directly transformed onto *E. coli* competent cells by heat shock at 42 °C for DNA amplification. Successful ligation is evaluated by screening bacterial colonies by PCR: a primer of the gene and a primer of the vector are used to ensure that the PCR amplifies only recombinant plasmids, as controlled on agarose gel (see **Figure 2.2**). Lastly, the colonies bearing these plasmids are cultured and a last PCR screening confirms the presence of the protein gene in the cultured cells. DNA sequencing is also performed as final check prior to transformation into *E. coli* strains for for protein expression.

Recombinant Protein Expression and Purification

Protein expression was performed in 1 L batch using 3L baffled culture flasks and induced with isopropyl β-d-1-thiogalactopyranoside (IPTG). IPTG binds to the lac repressor coded for in the plasmid (Lacl gene), which frees the lac operon in the plasmid, thus enabling recombinant protein expression. A 16 °C 'cold-shock' protocol is required for vector pCold TF – TEV to it ensure expression of target proteins at high yield and purity. Thanks to the cspA (cold-shock protein A) promoter, the target protein expression is upregulated on induction at low temperature while the expression of other proteins is inhibited. The expression of the Trigger Factor (TF) chaperone as a soluble tag enhances solubility and yields. Finally, addition of antibiotic ampicillin ensures the elimination of cells that don't acquire resistance from the plasmid (Amp^R gene), thus selecting for expression of the desired protein.

To release the target protein expressed in the cytoplasm, the bacterial calls are lysed via cell disruption using pressure. The target protein is found in the supernatant after centrifugation, which is filtered (0.45 µM) to remove particles prior to purification. To ease purification, the target protein fusion contains two features from the plasmid pCold TF –TEV: a Tobacco Etch Virus cleavage site (TEV) and a 6X histidine tag located at the N-terminal. The HisTag allows immobilized metal affinity chromatography (IMAC) to separate the fusion protein from other bacterial proteins. Immobilized protein can be eluted with a gradient of imidazole, which disrupts interactions of the proteins with the metal immobilized (Nickel here) on the column matrix. The TEV cleavage site allows the use of TEV protease to cleave the fusion containing the HisTag from the target protein. The latter is further purified by a second IMAC, in which the cleaved protein isn't immobilized but the cleaved tag and the uncleaved fusion are (see **Figure 2.2**).[154]

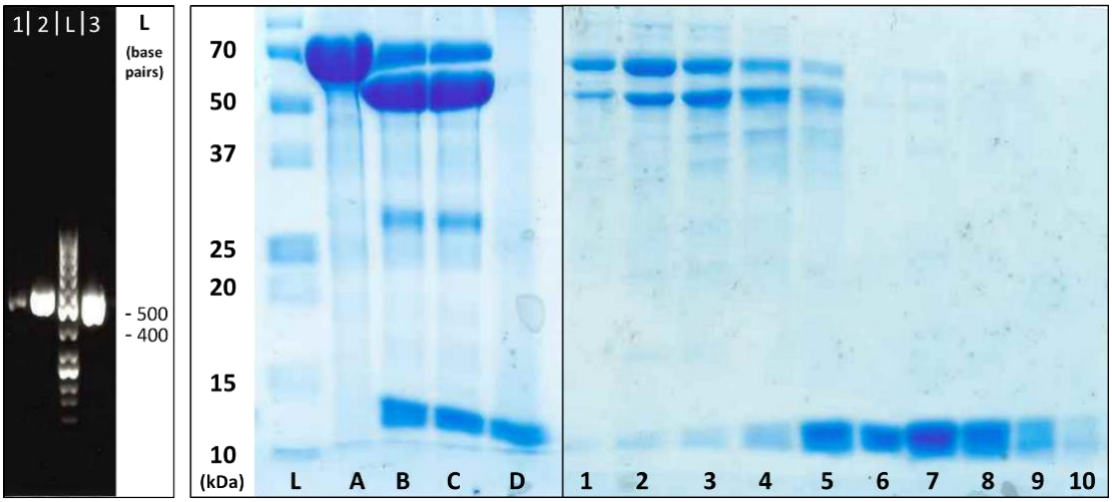


Figure 2.2. Left: Agar gel control of colonies: bands 2 and 3 show definitive presence of genetic material of the expected size: 518 bp (396 corresponding to the lectin gene and 122 to the vector). **Right:** 15% SDS-PAGE control of TEV protease-mediated cutting of the fusion (66 kDa) into the tag (52 kDa) and the lectin (14 kDa). Band A: pre-cutting, B and C: post-cutting (incomplete cutting), D: pure lectin reference. Bands 1-10 correspond to purification by SEC (bands 6-10: kept, bands 4-5: re-purified, bands 1-3: discarded).

A last step of purification involves Size Exclusion Chromatography (SEC): a porous matrix separates the lectin from the remaining protein contaminants by size (see **Figure 2.2**). In SEC, smaller particles interact more with the porous matrix, thus remaining a longer time in the column. By comparing their residence time against a pre-established standard, the size of the proteins can be roughly approximated to a molecular weight. This evaluation provides a first quality control of the recombinant lectin, since it is supposed to assemble as a trimer: it elutes as such from SEC. A second quality control is performed by electrophoresis: sodium dodecyl

sulfate–polyacrylamide gel electrophoresis (SDS-PAGE – 15%) allows to evaluate the purity of the lectin, but does not provide information on its multimeric assembly.

Finally, Dynamic Light Scattering (DLS) allows to accurately characterize the size of the lectin in solution. DLS harnesses the phenomenon of scattering of laser light by small particles dissolved in a sample. Due to the constant Brownian motion of the particles in suspension, the generated Rayleigh scattering fluctuates in intensity over time. By measuring this fluctuation of intensity, mathematical tools allow to physically characterize the particles. Among other parameters, the hydrodynamic radius of the particle is obtained and, in the case of proteins, can be correlated to a 'globular protein' molecular weight. Other useful information obtained through DLS is the monodispersity and homogeneity of the sample. Thus, DLS is a final quality control that ensures the purity, structural integrity, and multimeric assembly of the new protein construct.

Crystallization, X-ray Crystallography, and Structure Determination

Crystallography is a central tool for structural biology and provides unequivocal information about the structure of proteins and other molecules. Although the concepts that govern crystallography are deeply complex, the knowledge gained is invaluable and grants direct access to the behaviour of proteins in solid phase. The lack of access to information of dynamic processes is the other limitation of crystallography.

In order to obtain a valuable crystal structures, many steps are necessary. The first step is to crystallize the protein, usually achieved after screening many conditions and requiring homogeneous protein. The second is to perform X-ray diffraction on the crystal and collect quality data. The third is to use the appropriate mathematic and informatic tools to treat the data and eventually solve the crystal structure. The final step consists in validating and depositing the structure in the Protein Data Bank.[121]

Protein crystallization involves many variables and has a heavy empirical factor. The main factors influencing protein crystallization are concentration, temperature and crystallization conditions. Naturally, crystallization conditions include many other factors such as pH, precipitants and salts (in varying concentrations). It follows that obtaining crystals isn't simple, and many different conditions need to be screened to empirically deduct a 'trend' of preferred conditions. Furthermore, size and shape of the crystals depend on the conditions

of crystallization, meaning that obtaining a crystal doesn't necessarily imply the end of screening for conditions. What can be done to improve this process is to precisely regulate the reproducibility, especially in terms of protein purity, concentration and temperature.

Crystallization can be defined as the slow, controlled dehydration of protein. Nevertheless, to come out of solution doesn't instantly guarantee crystallization: precipitation occurs when protein doesn't organize itself before leaving its solvated state. As depicted in **Figure 2.3**, there is a fine line separating precipitation from crystallization. The precipitation zone is reached when the protein is too saturated and leaves the nucleation zone. In contrast, the nucleation zone allows crystals to appear without precipitating. Finally, the conditions leading to the metastable zone are ideal for growing existing crystals, but not for generating them. Since many parameters can control these evolutions, a number of slightly different diagrams exist for each pair of parameters. It also means that many strategies can be attempted. Following the idea of 'slow dehydration', one strategy is shown in **Figure 2.3**, which takes advantage of progressive dehydration with otherwise constant parameters.

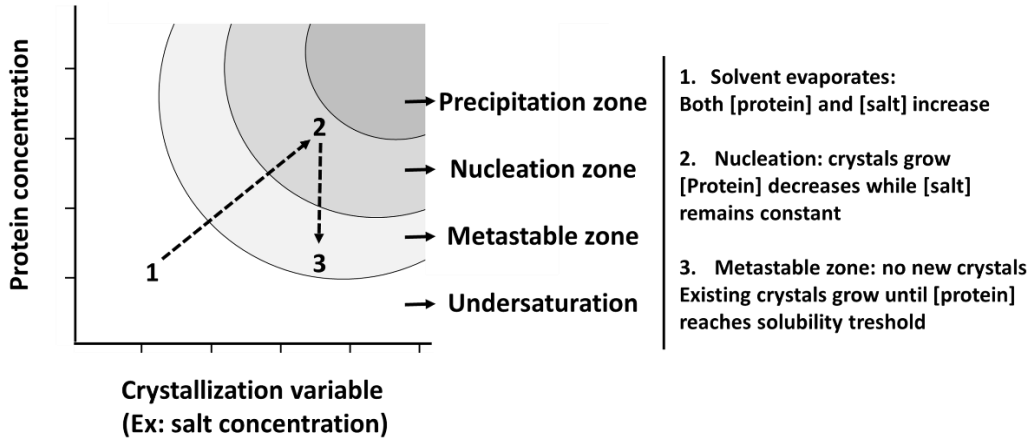


Figure 2.3. Solubility phase diagram for two variables. A simple example of crystallization strategy is described.

Practically speaking, a common way to maintain control over the many variables and increase reproducibility is to use purposefully developed crystallization techniques. One of such is the vapor diffusion technique, in which a drop of protein mixed with crystallization condition is sealed in a chamber holding crystallization condition in a reservoir (see **Figure 2.4**). The drop can either hang from a cover slip (*hanging drop*) or sit atop a small platform (*sitting drop*). By ensuring the solutions used are from the same batch and the chamber stays sealed at a constant temperature, precise control over most parameters can be achieved. At this level of control, screening is most effective and conditions for generating big and stable crystals can

likely be found. Nevertheless, upon changing a single parameter (for example: a mutation or a ligand for co-crystallization), the screening may need to start over.

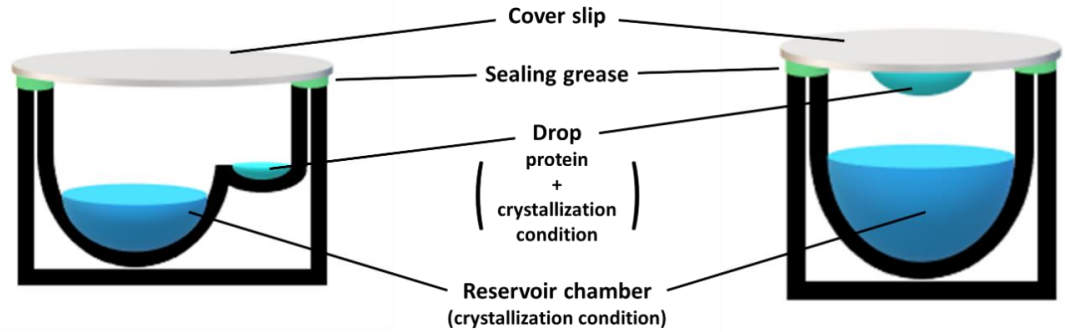


Figure 2.4. Sitting and hanging drop crystallization setups. The formation of crystals in the drop is assessed by microscopy.

In order to determine if the crystal obtained is suitable for diffraction experiments, it is can be transferred in a cryo-protectant solution in order to obtain vitreous ice when frozen in liquid nitrogen. This allows the crystal to be stored but also to limit the damages due to X-rays radiation.

Crystals are highly ordered structures in which a pattern of molecules repeats itself in a *lattice*. As such, a crystal lattice can be defined as an array of points in 3D space which have identical environments: the (arbitrarily positioned) *lattice points*. The crystal lattice can be defined by the 3D vectors that transform a lattice point into its identical mate. These vectors, along the angles separating them, define the minimal repeating unit of the lattice, which generates the whole crystal structure by simple translations: the *unit cell*. Its parameters are the vector lengths and the corresponding angles: a, b, c and α, β, γ. The possible characteristics of unit cells are finite, meaning crystals can be classified only into the 14 known categories called *Bravais lattices* (**Figure 2.5**).

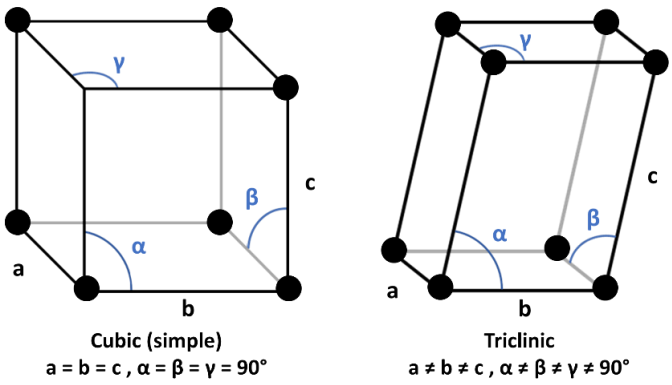


Figure 2.5. Examples of two Bravais lattices, the black circles correspond to lattice points.

On the other hand, there is more to crystalline order than translations: highly ordered crystals can also present symmetry by rotations, reflections and inversions. Indeed, a unit cell may contain elements that cannot be replicated by translation but yes by symmetry. Reducing the unit cell to the bare components that cannot be generated by symmetry produces the so-called *asymmetric unit*. By definition, the asymmetric unit can generate the unit cell by symmetry and the whole crystal by symmetry and translation. Accounting for these operations, the amount of different types of crystalline construction is 230, these are called *space groups*. Nevertheless, some symmetry operations are incompatible with the inherent chiral construction of proteins: for example, mirror planes would generate unnatural D-amino acids from their natural counterpart, which isn't observed in nature. By removing the groups associated with the incompatible symmetries, the final number of space groups pertinent to protein crystallography is 65. By understanding how a crystal is constituted, it becomes easier to collect the appropriate diffraction data to solve the protein structure.

The data collection is performed by bombarding the crystal with X-rays and recording the diffraction pattern obtained. The principle of this process follows the laws of optics: upon hitting the crystal lattice the incident ray is scattered. The diffraction pattern observed can be acquired in 2D by a detector: it shows as a number of spots - *reflections* - with varying intensities. For a given incidence angle of the ray, one 2D diffraction pattern is recorded. In it, each particular reflection corresponds to the diffraction of the beam on a different 'family of planes' of the crystal lattice. Since each family of planes is different and dissects the unit cell differently, it can be defined by the parameters that characterize that dissection (h, k, l). It follows that each reflection on the diffraction pattern is associated with the same coordinates, called *Miller indices* (see **Figure 2.6**). Another way to look at it is to define the *structure factor*: F_{hkl} is the mathematical description of how an incident beam is diffracted by a plane (h, k, l). It contains two parameters: the amplitude and the phase of the diffracted wave ($|F_{hkl}|$ and ϕ_{hkl}). For example, reflections are related to the lattice planes through *amplitudes*: higher intensity I_{hkl} corresponds to higher amplitude $|F_{hkl}|$ of the diffracted wave, which is proportional to the amount of matter contained in the plane that was hit. For example, a high-intensity reflection corresponds to a plane that features a relatively high number of electrons. Naturally, to gain as much information is necessary, many planes and many diffraction patterns need to be recorded: this is the reason for the rotation of the crystal during data collection. Depending on the amount of symmetry of a crystal, more or less data

needs to be gathered to solve its structure. Therefore, knowing the geometric parameters of crystals, in particular Bravais lattice, allows more effective data collection.

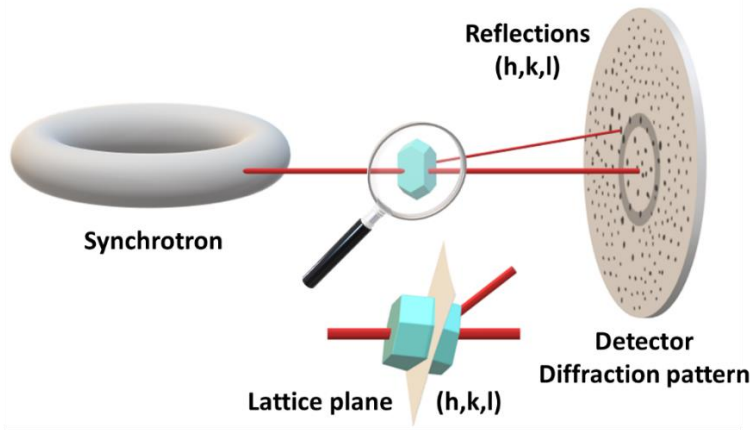


Figure 2.6. Schematic depiction of data collection through X-ray diffraction at a synchrotron. A reflection spot is related to the plane on which the ray diffracts and is characterized by the same Miller indices.

Once the diffraction data are collected, it has to be interpreted. For this, it is important to understand the relationship that links the position of atoms in the unit cell and the reflections obtained through data collection. As stated above, the intensity of a reflection I_{hkl} is related to the amplitude $|F_{hkl}|$ of the wave, which is, in turn, related to the atoms contained in the related plane in the crystal structure. To deduct the atomic spatial information from the intensities recorded, a mathematical tool is necessary: the *reciprocal space*. Reciprocal space is a 3D mathematical construct obtained by the inversion of real 3D space. The coordinates in this space are defined by the Miller indices (h, k, l), meaning that each point represents a reflection and its given intensity. It follows that intensities can be 'mapped out' in reciprocal space, roughly indicating which planes in the unit cell contain more or electrons. This part of the process is achieved thanks to the software included in the XDS package and the CCP4 program suite.[155-156] Although the electron density can be accessed through the mathematical relation between reflections, intensity and amplitude, crystal resolution is impossible before completing the dataset with the phases (angles of diffraction).

Indeed, obtaining the *phase* ϕ_{hkl} is not guaranteed by collecting diffraction patterns: the diffracting wave's amplitude can be obtained from the diffraction pattern, but not its angle of diffraction. 'Phasing' is the process of recovering the phases information and adding them to the observed data. There are many methods to do this: for example, MAD (multiple wavelength anomalous diffraction), which harnesses the special properties of diffraction on heavy atoms and uses them as reference marks in the mapping. A method enabled by the

growing number of solved structures is Molecular Replacement (MR), in which a homologous structure is used as reference model to orient the new structure and obtain its phases. These methods rely on the mathematical tool called *Patterson function*, which allows to use the collected amplitudes to map the proximity of atoms to each other in 3D space: a Patterson map. MR works by iterating translations and rotations of the data map until it matches the map generated by the model supplied.

Once the amplitudes and phases are obtained, electron density can be calculated on an initial atomic model. The last step of structure resolution is called 'refinement' and consists in iteratively adjusting the generated atomic model so that the positioned atoms match the electron density as well as possible and that the observed and calculated data correlate well. It also allows to add molecules that usually aren't accounted for, such as solvent, chemicals and ligands. Once the crystal structure is refined, it needs to be validated against geometry rules and quality fits before its release to the scientific community. The Protein Data Bank (PDB) manages platforms for validation and deposition of crystal structures: if no errors are spotted on the newly solved crystal structure, it is published.

2.2. Biophysical evaluation of lectins and their interactions

Isothermal Titration Calorimetry

ITC is a thermodynamic technique that allows characterization of protein/ligand interactions in terms of affinity, stoichiometry and thermodynamic parameters. Being a measure in solution, it represents the reality of an interaction more closely than surface-bound measures. As its name indicates, ITC is based on calorimetric measures along a titration, all while maintaining a constant temperature. In the usual case the titrant is the ligand, but the contrary experiment is also possible. Thus, the heat produced or required by the protein/ligand interaction is quantifiable.

To satisfy its principle, ITC requires accurate volumetric dosing, precise calorimetric measure, and fine control over temperature. These requirements are achieved by a modern calorimeter, in which temperature is kept constant in an adiabatic jacket that contains two cells. Of these two cells, one acts as temperature reference and is filled with water. The

second cell houses the sample and the thermodynamic variations ensuing from the injection of increasing quantities of titrant (see **Figure 2.7**).

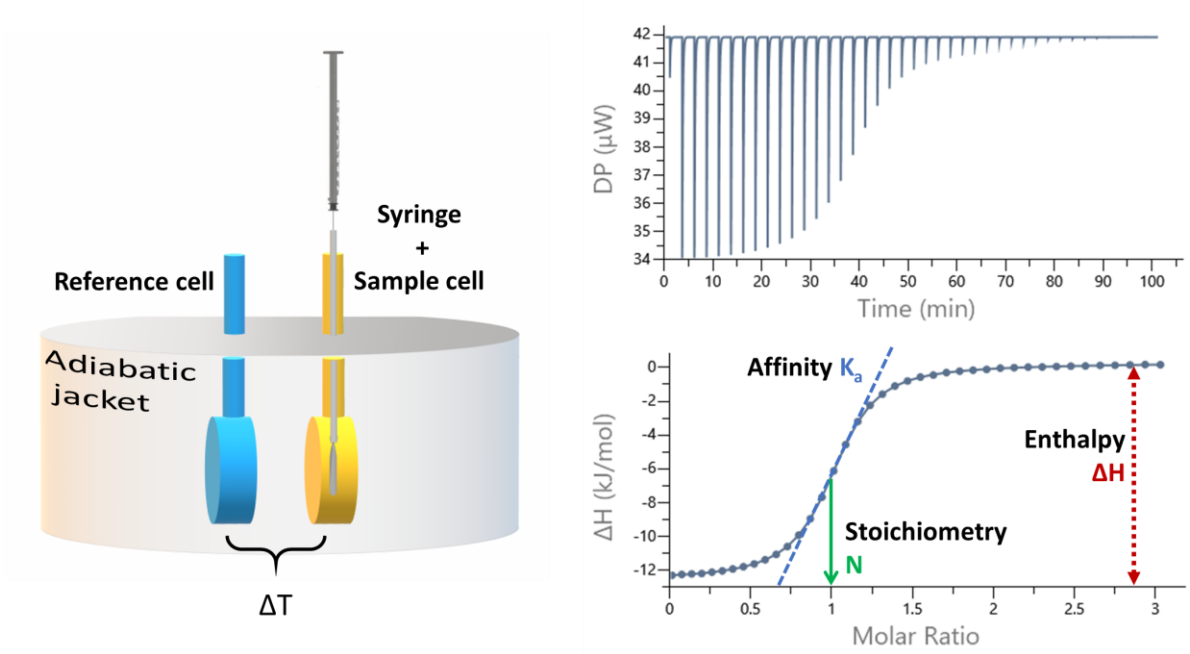


Figure 2.7. Left: Schematic microcalorimeter. **Right:** ITC annotated thermogram and titration curve.

These thermodynamic shifts in the sample cell translate into temperature deviations from the isotherm. To ensure the return to the isotherm, the apparatus provides more (or less) electric power to the system. This energy input is recorded as calories/second or watts, and related to the titration's advancement, in units of time. Once the thermogram peaks are integrated, the enthalpy variation for each injection can be plotted against the molar ratio. The molar ratio is calculated for the effective concentrations in the cell after each injection.[157]

In a typical experiment, a known concentration of ligand is injected in known volumes into a known volume of a known concentration of protein. The titration proceeds until the signal recorded for each injection is negligible, signifying saturation of the protein. The desired parameters are then mathematically obtained by fitting the titration curve. The height of the initial peaks is related to the variation of enthalpy (ΔH), the inflexion point corresponds to the stoichiometry and its slope to the affinity constant (see **Figure 2.7**). By relating the two expressions of the Gibbs free energy (**Equation 2.1**), the variation of entropy (ΔS) can be calculated, completing the thermodynamic picture of the interaction. Finally, the dissociation constant (K_D) can be directly obtained as the inverse of the binding affinity constant.

Equation 2.1: $-R \times T \times \ln K_a = \Delta G = \Delta H - T \times \Delta S$

R:	molar gas constant (8.314 $J.mol^{-1}.K^{-1}$)	ΔG:	Gibbs free energy ($J.mol^{-1}$)
T:	isothermal temperature (K)	ΔH:	enthalpy variation ($J.mol^{-1}$)
K_a:	binding affinity constant (M^{-1})	ΔS:	entropy variation ($J.mol^{-1}.K^{-1}$)

In order to obtain the most reliable data, blank thermograms can be subtracted to the experiment: titrations performed on a blank sample cell filled with buffer. Thanks to this, the heat produced by the buffer mismatch (dilution of the titrant from syringe to cell) can be separated from the interaction itself. Similarly, the reliability of the results increases when the shape of the fitted curve is an ideal sigmoid with a well-defined slope and inflexion point. This can be controlled by tuning the experimental factors such as concentrations and injection volume. The so-called 'c value' (**Equation 2.2**) is useful in this case:

Equation 2.2: $c = N \times K_a \times [protein]$

R:	'c value' (dimensionless)	K_a:	binding affinity constant (M^{-1})
N:	Stoichiometry (dimensionless)	[protein]:	sample cell concentration (M)

As a rule of thumb, c values are acceptable in the [1-1000] range and advised within [10-100].[158] In other words: for a given K_D value, the sample cell concentration should be comprised between 10 or 100 times that value, for a stoichiometry N = 1. A low c value will result in a flattened sigmoid and a high c value might miss the slope altogether by a sudden 'step'. To prevent this, the syringe concentration can be lowered for high c values: another rule of thumb is to aim for a [syringe]:[cell] ratio of 10 and modulate it [5 - 20] according to c. This ratio also allows for enough heat to be released in each injection, insuring the validity of the measures over the limit of detection. Finally, as transpires from above, low affinity interactions are difficult to study: affinities in the millimolar range (K_D = 1000 M) would require unsustainably high protein concentrations in the cell (1 mM and upwards). For these cases, a different experimental setup can be of use: a 'low c value' experiment.

'Low c-value' experiments presuppose low values for K_a and [protein], leading to c values below 1, hence the name (see **Equation 2.2**). In these cases it's not advised to keep a [syringe]:[cell] ratio of 10, which would require many injections to saturate the protein with ligand, exceeding the volumetric capacity of the machine. Instead, a much higher ratio is advised, to attempt a final free ligand concentration in the range [10 x K_D - 100 x K_D], corresponding to 90% and 99% saturation, respectively. It's worth to note that the high ligand concentrations required can conflict with its solubility. Naturally, at these concentrations,

only the upper part of the sigmoid will be acquired, meaning that the variation of enthalpy is not available for calculation. On the other hand, saturation guarantees the K_a value, provided the stoichiometry N is known and fixed to its value during fitting. Consequently, appraisal of the affinity is achievable under specific circumstances for lower affinity ligands, but thermodynamic study is inaccessible or unreliable.

Surface Plasmon Resonance

SPR experiments characterize the interaction of a ligand in solution with a surface-bound partner. As such, experimentation is limited by the necessity to choose and tether one partner of the interaction. On the other hand, surface-bound interactions are interesting to study in the context of lectins, as they usually interact with their ligands present on biological surfaces and interfaces. The effects of multivalency and epitope presentation may be better observed in such a setting.

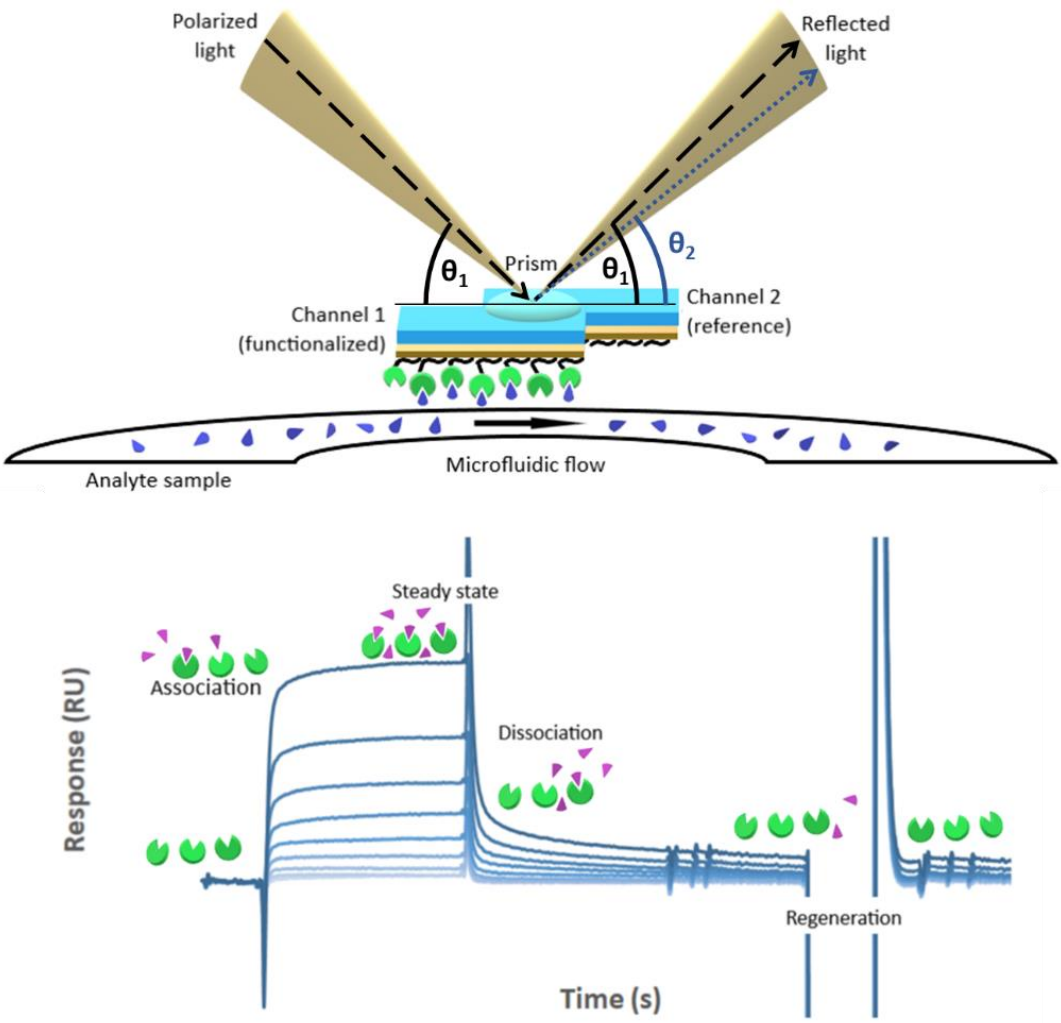


Figure 2.8. Top: Schematic SPR apparatus. **Bottom:** Sensogram depicting the different steps of a run.

The principle of SPR resides in the physical phenomenon called surface plasmon polaritons (SPPs). SPPs are electromagnetic waves generated when energy is reflected on a metal/dielectric interface.[159] For example, when monochromatic polarized light is reflected on a gold-coated SPR chip, not only it reflects, but it also penetrates the material in the form of resonant oscillation of electrons. Naturally, the extent to this effect is highly dependent on the physical characteristics of the surface: on slightly different surfaces, light will be reflected differently and cast different angles of reflection and absorption (see **Figure 2.8**).

The so-called resonance angle θ corresponds to the angle at which the reflected light has the lowest intensity (highest resonance at the surface). By measuring and recording this angle, SPR aims to characterize the nanometric changes happening at the aforementioned surface, for example, chemical functionalization.

In a classical SPR experiment, a macromolecule such as a protein is covalently bound on one channel of the chip surface while a microfluidics system streams the ligand sample along the chip. A second channel is kept blank for reference. During the association time, the ligand flows onto the surface and interacts with the protein. This interaction modifies the surface environment, shifting the characteristic resonance angle θ. During the dissociation, ligand flow stops and is replaced by buffer, progressively washing out the ligand. Lastly, a regenerating step using buffer or a regeneration solution recovers the initial surface state and angle. The changes in refractive index and resonance angle are measured as a response (R) in 'Resonance Units' – RU, which are related to the variation in area density (1000 RU $\leftrightarrow$ $\Delta\theta$ = 0.1 ° $\leftrightarrow$ $\Delta\rho_A$ = 1 ng.mm^{-2}). A sensogram describes the experiment by plotting response against time. By varying the concentrations of ligand used in different runs and analysing the association/dissociation profiles, measures of affinity and kinetics can be obtained through mathematical fitting of the data (see **Figure 2.8**).

Among the possible experiments available from SPR, the measure of binding affinity takes advantage of the steady state. Indeed, when a steady state is attained, the number of molecules dissociating from the interaction is equal to the number of molecules associating: the rate of complex formation is zero. The rate of complex [PL] formation can be defined as the difference between associating (free) molecules and dissociating (bound) molecules and is proportional to the measured response R (see **Equation 2.3** and **2.4**).

Equation 2.3:

$$\frac{d[PL]}{dt} = k_a \times [L]_{free} \times [P]_{free} - k_d \times [PL] = k_a \times [L] \times ([P]_{tot} - [PL]) - k_d \times [PL]$$

[PL]:	Complex concentration (M)	k_a:	Association rate constant ($M^{-1}.s^{-1}$)
$[L]_{free}$ = [L]:	Free ligand concentration (M)	k_d:	Dissociation rate constant (s^{-1})
$[P]_{free}$:	Free protein concentration (M)	t:	Time (s)

$[P]_{tot}$ and [PL] are proportional to the response

Equation 2.4: $\frac{dR}{dt} = k_a \times [L] \times (R_{max} - R_{eq}) - k_d \times R_{eq} = 0$

R:	Response (RU)	k_a:	Association rate constant ($M^{-1}.s^{-1}$)
R_{max}:	Maximal response (theoretical saturation)	R_{eq}:	Response at steady state

From here, we can extract a new expression (**Equation 2.5**) that relates the response observed at steady state (R_{eq}) for a given concentration of ligand ([L]) with the quotient of association rates, also known as dissociation constant (K_D).

Equation 2.5: $R_{eq} = \frac{[L] \times R_{max}}{[L] + K_D} + R_{off}$

R_{eq}:	Response at steady state (RU)	$K_D = k_d/k_a$:	Dissociation constant (M)
R_{max}:	Maximal response (RU)	R_{off}:	Offset parameter (RU)

Different runs at different concentrations will provide enough data to fit **Equation 2.5**, providing the values for theoretical R_{max}, K_D, and an offset parameter R_{off}. Although unnecessary, for reaching R_{max} in practice, steady state is not enough: the chip sites need to be saturated, meaning the concentration of ligand [L] in flow must be over the 100 x K_D threshold. Noteworthy, R_{eq} observed at steady state equals half of R_{max} for a concentration of ligand equal to K_D (from **Equation 2.5**). Therefore, a good 'steady state affinity' experiment features multiple runs at ligand concentrations in the range [0.1 x K_D – 20 x K_D]. A blank (zero concentration) run allows a second reference subtraction, for more reliable data. As it transpires, affinities in the micromolar range are difficult to measure due to the high ligand concentrations needed.

Another possible SPR experiment is designed to unravel the kinetic data of the interaction: by studying the shape of association and dissociation curves, mathematical fitting can provide

values for the association and dissociation rate constants (k_a and k_d). Nevertheless, this kinetic model fitting is only adapted for a range of values: rates that exceed the limit cannot be fitted. On the other hand, if the kinetic study is successful, the dissociation constant (K_D) can be derived from the kinetic data. SPR can also be used as a screening tool for the binding of new molecules to a target-coated chip. Ligands can be distinguished from other molecules by evaluating the responses obtained for different samples. Regarding multivalency, a protein-coated chip can provide the K_D of an interaction, but a ligand-coated chip can provide information on avidity and the impact of multivalency. Lastly, SPR can allow to calculate IC_{50} values when the interaction of the surface with a ligand at a constant concentration is disrupted by increasing concentrations of competitor.[160]

Differential Scanning Calorimetry

DCS is a calorimetry technique that allows characterization of small thermal events in solution. For example, DSC allows precise study of the unfolding (denaturation) of a protein upon temperature gradient. This resembles other similar techniques: TSA (Thermal Shift Assay), which uses a dye to characterize unfolding events, or nanoDSF (Differential Scanning Fluorimetry), which uses the intrinsic fluorescence of aromatic residues. Contrary to these two, DSC is much more precise, as it uses calorimetry to quantify a thermal event: it can work on diluted samples and does not depend on additives or the amount of aromatic residues present.

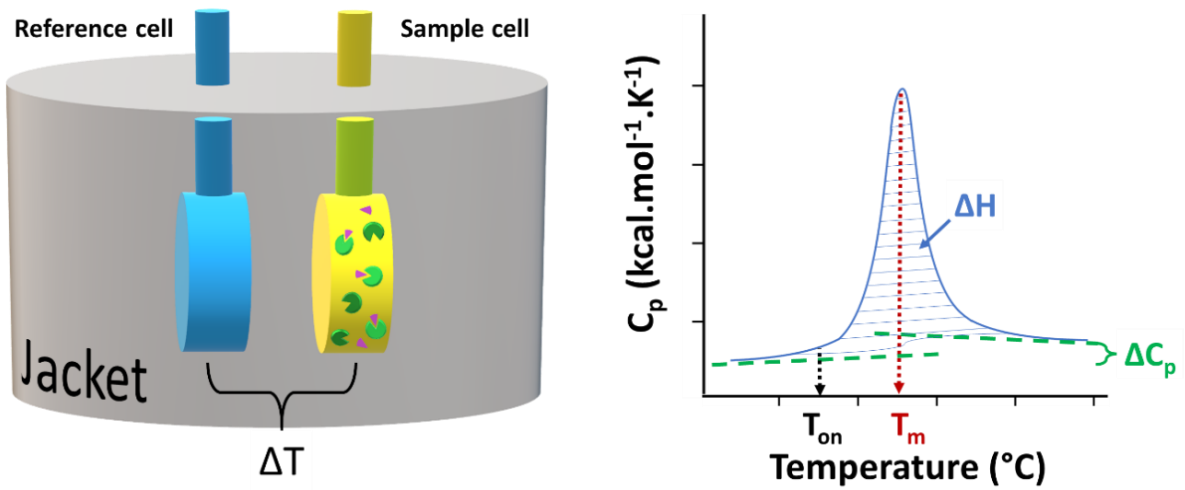


Figure 2.9. Left: Schematic DSC calorimeter. **Right:** DSC annotated thermogram of a thermal event.

The apparatus consists in a thermo-controlled jacket that houses two cells. Similar to ITC, DSC measures the heat changes in a protein-filled cell against a 'blank' (buffer-filled) cell. On the other hand, DSC isn't isothermal: it uses a temperature gradient or 'scan rate' to induce thermodynamic changes on the sample, while the cell is sealed at a constant pressure. By measuring the heat uptake caused by the thermal events and plotting it in a thermogram, this technique provides values for the following parameters: onset temperature (T_{On}), transition midpoint (T_m), enthalpy (ΔH) and heat capacity (ΔC_p) of a given event (**Figure 2.9**).

One of the main uses of DSC is to characterize the thermal stability of proteins and track its evolution across different conditions and modifications. Going further, the sensitivity of DSC allows comparison between protein samples and samples containing also ligands. By observing the thermograms, the shift in transition midpoint(s) can provide information on binding interactions and their effect on thermal stability. Thanks to this thermal shift, usually observed to be positive for stabilized complexes, DSC can be used to screen and rank molecules by their capacity to bind and stabilize the protein.

Saturation Transfer Difference NMR

Saturation Transfer Difference (STD-NMR) is a ligand-based NMR technique used to study the interaction between a protein and small ligands. In the STD experiment, a macromolecule is analysed in presence of a large excess of a small molecule ligand. Firstly, the protein is selectively saturated by a radiofrequency, then the magnetization is transferred via nuclear Overhauser effect (NOE) to spatially close protons of bound ligands. When the ligand returns to solution and its signal is acquired, the strongest intensity is observed for the protons in closest proximity to the protein (see **Figure 2.10**).[161]

Therefore, this technique can be used to describe protein/ligand interactions and structurally map the epitopes recognized. Furthermore, with strong-enough signals, the epitope mapping can be quantified in relation to the 'degree of saturation', meaning the normalized percentage of signal for each proton. With this, the protons (and moieties) essential to the interaction can be exposed. Since the complex is dynamic in solution, it is complementary to crystallography and can provide information on transient events.

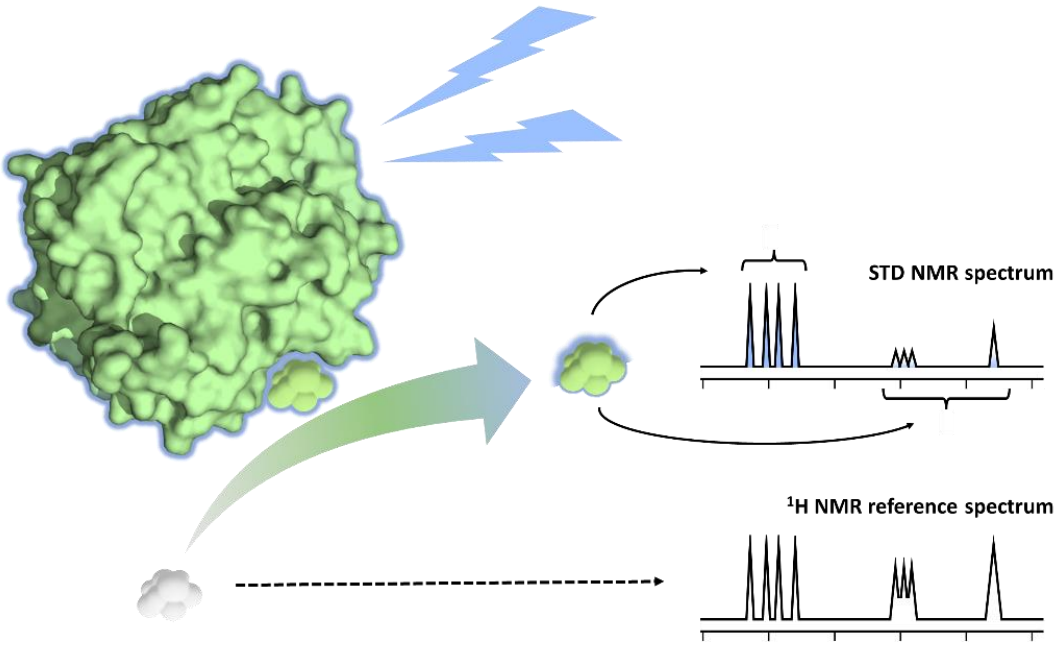


Figure 2.10. STD-NMR experiment. A protein is selectively irradiated (blue). The saturation is transferred to its ligand upon binding. The protons of the released ligand retain irradiation proportional to their proximity to the protein. The differences between STD and reference spectra allow to map the interaction.

Another useful application of STD-NMR is its ability to screen structures for binding, even for cocktails of molecules: the STD spectra only shows the signals of the binding molecules, separating the hits from other non-binding structures. The only requisite for this is accurate ^{1}H NMR characterization of the molecules and wariness of overlapping signals.

3. SCOPE OF THE THESIS

3.1. Limits to previous studies

As briefly stated in the introduction, there were two limiting factors for the thorough study of BC2L-C, in particular its N-terminal domain and its interactions with fucoside ligands:

The first limitation related to the construct originally prepared for study: BC2L-C-nt was 187 residues-long, composed of the N-terminal lectin domain (130 residues), the flexible segment (26 residues) and the uncleavable HisTag (31 residues).[134] The flexible character of the C-terminal extremity, which was only known in retrospective, was detrimental to protein stability: aggregation was common and crystallization with oligosaccharide ligands wasn't achieved. A second-generation construct wouldn't need to feature the C-terminal HisTag or the 31 residues-long linker region.

The second limitation was the lack of a reliable framework to develop high-affinity ligands for the N-terminal domain. Potential ligands were evaluated by their capacity to inhibit the hemagglutination of red blood cells. Although this is a viable functional evaluation, rational design and enhancement of ligands requires a closer look at the interaction in question. Both structural and biophysical methods can provide critical information in regard to protein/ligand interactions and measures of affinity.

3.2. The PhD4GlycoDrug Consortium

Providing the framework for this thesis, the PhD4GlycoDrug consortium is a 'Marie Skłodowska-Curie action' and an 'Innovative Training Network' (MSC - ITN): a European project funded by the Horizon 2020 research programme.[162] The main objective of PhD4GlycoDrug is to train a new generation of glyco-scientists and to pursue the development of carbohydrate-based therapeutic molecules (*glycodrugs*) targeted to different areas. Among the different topics addressed by this consortium, structural biology, organic synthesis and medicinal chemistry are of particular help for characterizing and antagonizing pathogenic lectins.

A second role of this consortium is to establish a network that fosters collaboration and partnership. Being a European Joint Doctorate, PhD4GlycoDrug allows PhD students to be

enrolled in two universities and combine their unique domains of expertise. As such, two complementary skillsets could be combined for this project: the expertise in lectin characterization of the Structural and Molecular Glycobiology (GMBS) group of CERMAV (University of Grenoble) and the expertise in glycomimetic synthesis of the Bernardi group at the University of Milan. On a related note, collaborations between PhD students are encouraged in the consortium. Complementary skillsets allow two or more PhD students to contribute to the same project from different angles in a convergent manner. For example, the work presented in this thesis is complemented by the computational work of Kanhaya Lal for the design of lectin antagonists through molecular modelling.

Lastly, a third role of the consortium is to disseminate free scientific information to the scientific community and the general public, particularly regarding carbohydrates and their therapeutic applications. This objective was implemented in partnership with the Glycopedia platform, which has a similar alignment in regard to public dissemination of glycoscience.[163] Communications for the benefit for the scientific community were produced under this partnership, such as the review article presented in the **APPENDIX 8.1: Scientific Communication: secondment at Glycopedia**.[164]

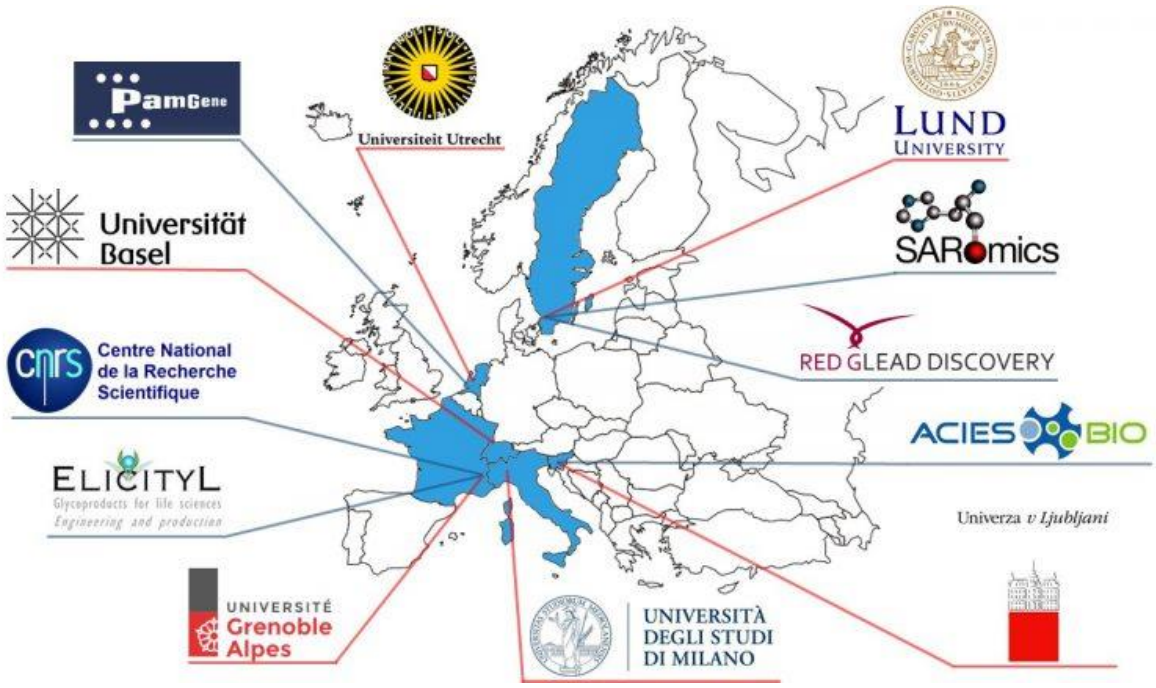


Figure 3.1. The PhD4GlycoDrug consortium.

3.3. Thesis Objective

It has been established that MDR *B. cenocepacia* represents a contemporary threat that is being addressed through many channels. One of such channels is anti-adhesion therapy (AAT), which aims to disrupt the virulence mechanisms of the bacterium. Considering the current successes of targeting virulent lectins of other pathogens, inhibition of *B. cenocepacia*'s lectin family BC2L holds promise for AAT. In particular, the superlectin BC2L-C is an interesting target, owing to the relation between its C-terminus and known virulence factors, and the novelty of its N-terminal lectin domain. Since BC2L-C-N_{ter} remains relatively uncharted as a target, it becomes the logical choice to pursue a research campaign. Thus, learning more about this lectin and how to antagonize it is a relevant step in the larger endeavour against *B. cenocepacia*.

As the name of this project indicates, the 'Design, Synthesis and Evaluation of antagonists towards BC2L-C' has three main objectives:

- The production and characterization of a new construct for BC2L-C-N_{ter}
- The rational design and synthesis of fucoside antagonists for this target
- The evaluation of the produced glycomimetics against their target

The first part of the project will provide critical information about the target and its interactions to enable rational design of ligands. The second part will be complemented by computational modelling to produce a first generation of BC2L-C-N_{ter} antagonists. Moreover, it will establish a framework for the synthesis of future generations of ligands. The third and last part will determine the validity of the previous steps and provide a reliable system to evaluate ligands, which will be useful not only to quantify the success of the project but also to serve all subsequent attempts to engage this target.

4. A NEW CONSTRUCT FOR BC2L-C-N_{TER}

4.1. Summary

The first nine months of the thesis were spent in the Structural and Molecular Glycobiology (GMBS) group of CERMAV in 2018. During this time, a new recombinant construct was prepared for the lectin domain BC2L-C-N_{ter}: rBC2LCN. It was 130 residues-long and proved to be stable and generally good to handle, characterize and use for the biophysical evaluation of its interactions. Its lectin function wasn't impaired, showing that the residues removed were not essential. It was characterized in terms of weight and size, then its binding to oligosaccharides was evaluated by ITC. It led to unprecedented co-crystallization with oligosaccharide ligands (H-type 1 and Globo H/H-type 3). These two crystal structures and the article describing the work were published in early 2020.[165]

After a 14-month stay in the University of Milan, work at GBMS resumed in early 2020, partly aiming to supplement the aforementioned work. Among the newer achievements, the ITC characterization was completed with carbohydrate ligands of varying sizes, identifying the best ligand known to date: the H-type 1 trisaccharide (K_D = 25 µM). Additionally, a new crystal structure was obtained for the complex BC2L-C-N_{ter}/Lewis y, revealing new information about the binding interaction.

The following sub-sections present the 2020 article and the further knowledge gained in 2021. Taken together the successful new construct and the information gained open the gate to a better understanding of the function of BC2L-C, its interactions, and how to antagonize them.

4.2. Article: BC2L-C N-Terminal Lectin Domain Complexed with Histo Blood Group Oligosaccharides Provides New Structural Information

molecules

MDPI

Article

BC2L-C N-Terminal Lectin Domain Complexed with Histo Blood Group Oligosaccharides Provides New Structural Information

Rafael Bermeo [1,2], Anna Bernardi [2] and Annabelle Varrot [1,*]

1 Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France; rafael.bermeo@cermav.cnrs.fr
2 Università degli Studi di Milano, Dip. Chimica, via Golgi 19, 20133 Milano, Italy; anna.bernardi@unimi.it
* Correspondence: annabelle.varrot@cermav.cnrs.fr; Tel.: +33-476037634

Academic Editor: Derek J. McPhee
Received: 7 December 2019; Accepted: 29 December 2019; Published: 7 January 2020

Abstract: Lectins mediate adhesion of pathogens to host tissues, filling in a key role in the first steps of infection. Belonging to the opportunistic pathogen *Burkholderia cenocepacia*, BC2L-C is a superlectin with dual carbohydrate specificity, believed to mediate cross-linking between bacteria and host cells. Its C-terminal domain binds to bacterial mannosides while its N-terminal domain (BCL2-CN) recognizes fucosylated human epitopes. BC2L-CN presents a tumor necrosis factor alpha (TNF-α) fold previously unseen in lectins with a novel fucose binding mode. We report, here, the production of a novel recombinant form of BC2L-CN (rBC2L-CN2), which allowed better protein stability and unprecedented co-crystallization with oligosaccharides. Isothermal calorimetry measurements showed no detrimental effect on ligand binding and data were obtained on the binding of Globo H hexasaccharide and L-galactose. Crystal structures of rBC2L-CN2 were solved in complex with two blood group antigens: H-type 1 and H-type 3 (Globo H) by X-ray crystallography. They provide new structural information on the binding site, of importance for the structural-based design of glycodrugs as new antimicrobials with antiadhesive properties.

Keywords: TNF-like lectin; fucosides; blood group antigen; crystallography

1. Introduction

Nosocomial infections, also known as healthcare-associated infections (HCAIs), have always been part of the public health scene. They represent an alarming problem due to the rise of multidrug-resistant (MDR) pathogens. They are correlated with increased hospitalization, morbidity, mortality, and financial burden. *Burkholderia cenocepacia* is one of the notorious bacteria associated with this issue. This opportunistic pathogen is not only a widespread Gram-negative and biofilm-forming bacterium, but also a MDR bacterium with intrinsic resistance to multiple classes of antibiotics [1,2]. It represents a major hazard to hospitalized patients immunocompromised, affected with cystic fibrosis, or critically ill in the intensive care units [3,4]. *B. cenocepacia* belongs to the *Burkholderia cepacia* complex (BCC), a compendium of human pathogens from the *Burkholderia* genus that can potentially elicit the "cepacia syndrome" in patients, leading to a rapid decline of pulmonary function often leading to fatal outcome [5,6]. *B. cenocepacia* is one of the BCC species most frequently isolated from patients.

As other opportunistic bacteria, *B. cenocepacia* employs carbohydrate binding proteins (i.e., lectins) as virulence factors to target host tissues through recognition and subsequent adhesion to the glycoconjugates on the cell surface [7]. Among the different lectins produced by this bacterium, BC2L-C (272 amino acids) stands out to our interest. It was initially identified due to the strong homology of its C-terminal to LecB, a fucose binding lectin belonging to *Pseudomonas aeruginosa*, an even more notorious opportunistic bacterium involved in HCAIs [8]. BC2L-C also presented at its N-terminus a domain

of unknown function, which was later revealed to be also a lectin domain (BC2L-CN). Establishing a new lectin family, BC2L-CN showed specificity for fucosides, especially blood group antigens, for which the measured affinity reached micromolar range. Blood group antigens are commonly targeted by lectins during microbial infections since they are highly present on the surface of epithelial cells, mucosa, and mucus [9]. BC2L-CN was the first lectin to display a trimeric arrangement of the jelly roll fold primarily observed in tumor necrosis factor α (TNF-α). A novel binding site and binding mode for fucosides was also uncovered at each interface between two protomers, which implicated an essential arginine residue [10]. Furthermore, BC2L-C was the first superlectin to be characterized with its two lectin domains assembling to form a final hexameric structure [11].

Lectins play a key role in microbial adhesion to host cells and between bacterial cells during biofilm formation, hence, antiadhesive glycodrugs or glycomimetics inhibiting lectins are attracting increasing attention as novel antimicrobials [12,13]. In this context, and due to its affinity for human oligosaccharides, BC2L-CN becomes a relevant target. Nevertheless, prior to any probing, additional information is required on its interactions with oligosaccharides. The original construct to produce BC2L-CN in recombinant form allowed crystallization only with the α-methyl fucoside. It consisted of the first 156 residues of BC2L-C with an additional C-terminal tag of 31 amino acids [10]. Structural analysis revealed that only the first 130 residues constitute the N-terminal domain of BC2L-C. The remaining residues belong to the linker between the two lectin domains and tag. These residues could not be observed in the crystal structure, hinting at high flexibility, which would explain the relative instability of the purified protein. Likewise, the impossibility to crystallize it with oligosaccharides could be attributed to this structural shortcoming, as those residues could hinder oligosaccharide binding. In order to overcome these limitations and expand the structural study of the protein, a shorter construct was designed comprising only the first 132 residues of BC2L-C and a cleavable N-terminal fusion to help expression and purification.

Here, we present the cloning, expression, and purification of the new construct of BC2L-CN (rBC2L-CN2) as well as the evaluation of its binding affinity for different fucosides by isothermal microcalorimetry (ITC). We also describe its structures in complex with H-type 1 and Globo H blood group antigens at 1.6 and 1.9 Å resolution, respectively, with details on the new interactions observed.

2. Results

2.1. New Construct for Recombinant of BC2L-CN

To improve upon the shortcomings of the previous construct of BC2L-CN on stability, flexibility, and difficulty to obtain crystal complexes with oligosaccharides, a new design was conceived. Based on the structural examination of first BC2L-CN structure (PDB 2WQ4), a construct featuring solely the ordered amino acids of the N-terminal domain was designed (1 to 131). The desired DNA sequence was amplified by PCR using the old construct as a template, purified, and subcloned into two vectors of expression: pET-TEV (Pet28a derivative) and pCold-TEV, using the restriction sites NdeI and XhoI. The pCold-TEV vector is a modified pCold-TF vector in which the factor Xa cleavage site (TATCGAAGGTAGG) was replaced with a TEV cleavage site (GAAAACCTGTATTTTCAGGGC) by mutagenesis using appropriate oligonucleotides [14–16]. The protein expressed by those vectors presented a N-terminal fusion that could be removed by proteolysis using the tobacco etch virus (TEV) protease and also included a 6-Histidin tag to ease purification. Successful expression was obtained overnight at 16 °C after induction with isopropyl-β-D-thiogalactoside (IPTG). Soluble protein was purified using immobilized nickel affinity chromatography thanks to the presence of the His-tag, over two steps: before and after TEV cleavage. While the TEV protease failed to cleave the 20 amino acids-long tag issued from pET-TEV, the larger fusion expressed using pCold-TEV allowed a conformation compatible with TEV cleavage, thus justifying its choice for further studies. It provided a 605 residues-long protein (66.4 kDa) with a N-terminal fusion (52.4 kDa) containing the trigger factor, a prokaryotic ribosome-associated chaperone (46 kDa), which aids co-translational folding of newly

expressed polypeptides (Figure 1A). Size exclusion chromatography (SEC) was performed as a final purification step, leading to a 95% pure protein with an average yield of 5.2 mg.L^{-1} of culture medium. In this form, concentrated protein (up to 15 mg.mL^{-1}–1.070 mM) was stable for several months both at 4 °C and −20 °C. Very little precipitation could be observed after one year storage in the fridge.

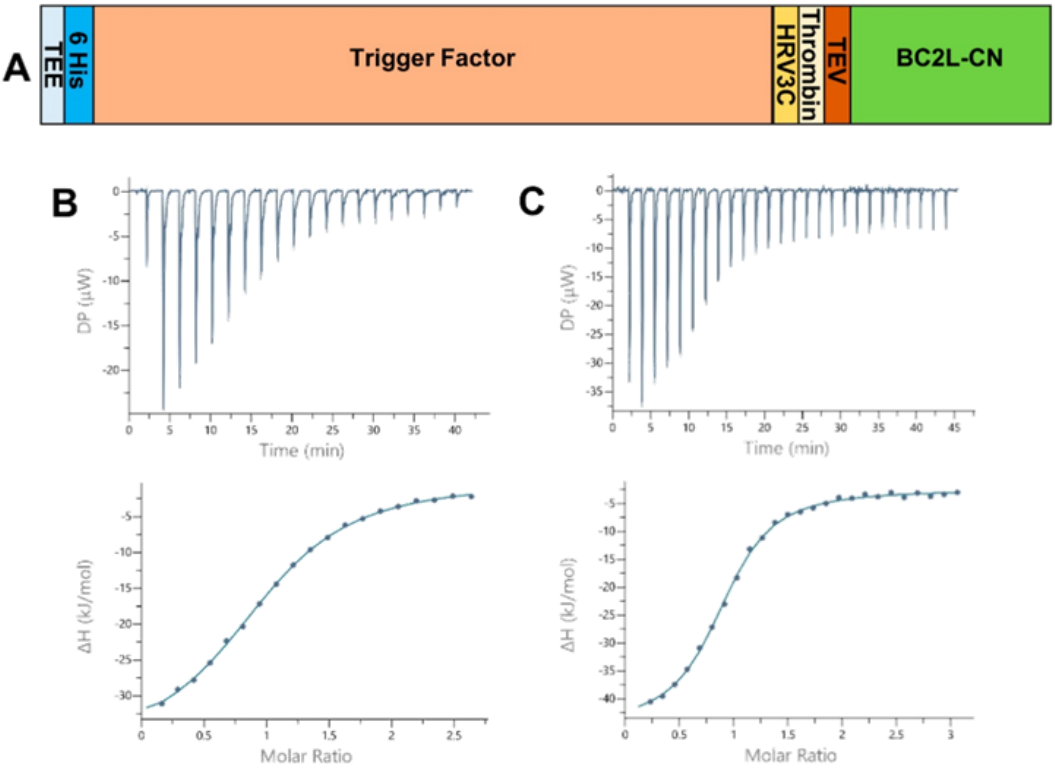


Figure 1. (**A**) Schematic of the expression construct of BC2L-CN2 in pCold-TEV. The N-terminal fusion presents a translation enhancing element (TEE), the trigger factor (TF) and 3 protease sites (human rhinovirus 3c (HRV 3c), thrombin, and tobacco etch virus (TEV)). The full construct leads to a protein of 605 and 134 amino acids before and after cleavage with the TEV protease. (**B**,**C**) Isothermal microcalorimetry data. Titration of rBC2L-CN2 (314–350 µM) by H-type 1 tetrasaccharide ((**B**), 4 mM) and Globo H hexasaccharide ((**C**), 10 mM) at 25 °C. The isothermal titration microcalorimetry (ITC) thermogram is displayed in the upper panel and the integration of data with curve fitted for "one binding site" model in the lower panel.

2.2. Physical Analysis of the rBC2L-CN2

Size exclusion chromatography on a methacrylate-based matrix (Enrich70, Bio-Rad, Marnes-la-Coquette, France) was used to remove the last contaminants, remaining uncleaved protein, and aggregates. It also allowed us to confirm the oligomerization and homogeneity of the protein. The molecular size of the sharp eluted peak corresponding to rBC2L-CN2 on SDS PAGE was approximated to 33 kDa by the means of a calibration curve (data not shown). This value was compatible with the expected trimer, which has an estimated molecular weight of 42 kDA. The weight discrepancy could be due to some nonspecific interactions of the lectin with the gel filtration matrix delaying the protein elution. This was observed before for other lectins when confronted to a saccharidic matrix [17].

To further confirm the trimeric arrangement, dynamic light scattering (DLS) measurements were performed at 22 °C on the purified protein. A monodisperse peak was obtained leading to a hydrodynamic radius of 2.94 nm which corresponded to a protein with a molecular weight of 41.9 kDa, matching the expected weight of the trimer (data not shown). The deletion of the 56 amino acids of the original construct did not have an impact on protein folding or on trimer formation of the N-terminal domain.

2.3. Affinity Analysis and Activity Validation

Isothermal titration calorimetry (ITC) measurements were then performed on rBC2L-CN2 to verify that its binding to carbohydrate was not modified. The affinity (dissociation constant) and thermodynamic contributions were obtained for the interactions with various human blood group antigens including H-type 1 tetrasaccharide (Fucα1-2Galβ1-3GlcNAcβ1-3Gal), Lewis Y pentasaccharide (Fucα1-2Galβ1-4(Fucα1-3)GlcNAc β1-3Gal), and Globo H hexasaccharide (Fucα1-2Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glc), which presents H-type 3 antigen. Due to its similarity to L-fucose, the monosaccharide L-galactose was also analyzed. This experiment was suggested by a close analysis of the Fuc methyl group binding region in the available X-ray structures. Indeed, the fucose methyl group moiety was found in lipophilic contact with the side chain of Tyr 48, and it was flanked by the hydroxyl group of Thr83, whose side chain lined the pocket on the opposite side. This suggested that a hydroxyl group at Fuc C6 (leading to the L-Gal structure) may actually be accepted by the lectin. Some of the ITC titration values measured were compared to those previously obtained with the original construct, also reported in Table 1.

Table 1. Affinity measurements for the binding of various carbohydrates to rBC2L-CN2 by isothermal microcalorimetry at 25 °C.

Ligand	n	K_d (μM)	−ΔG (kJ/mol)	−ΔH (kJ/mol)	−TΔS (kJ/mol)	Ref
Lewis Y	0.99 [a]	52.6	24.4	43.3	18.8	This study
	0.98 ± 0.03	53.9 ± 2.9	24.4 ± 0.2	34.9 ± 0.3	10.5	[10]
H-type 1	1.01 [a]	56.6	24.3	37.5	13.2	This study
	0.93 ± 0.02	77.2 ± 1.5	23.5 ± 0.2	23.0 ± 0.3	−0.5	[10]
L-galactose	1 [b]	2000	NA	NA	NA	This study
GloboH (H-type 3)	0.83 ± 0.06	26.05± 1.7	26.1 ± 0.2	46.1 ± 3.9	20.1	This study

NA: not applicable. Averages values and experimental errors are given when at least two independent measurements were made. [a] In the case of H-type 1 and Lewis Y, parameters result from unique measurements only, as they were in the same range as the previous construct. [b] Fixed during the fitting procedure.

All thermograms started with strongly exothermic peaks, characteristic of enthalpy-driven interactions, followed by peaks decreasing in height while saturation was achieved (Figure 1B,C). When strong binding affinity led to a sigmoidal curve, the stoichiometry of binding was determined with values close to 1.0, indicating one sugar binding site per monomer. For low-affinity ligands, such as the L-galactose monosaccharide, the stoichiometry was fixed to 1.0. We started by testing Lewis Y and H-type 1, which previously presented the best affinities for BC2L-CN. The values obtained after a single measurement were within the expected range, proving that rBC2L-CN2 was functional (Table 1 and Figure 1B). Despite a lack of reliable thermodynamic data, the affinity for L-galactose could be estimated in the millimolar range, with a K_d of 2 mM, which was similar to what was previously measured for alpha-methyl-L-fucoside (K_d = 2.7 mM) [10]. This supported our structure-based hypothesis that a hydroxyl substitution at fucose C6 would not be detrimental for binding and offers a useful starting point for the design of unnatural ligands of BC2L-C N-terminal domain. We also determined the thermodynamic parameters for the interaction of rBC2L-CN2 with Globo H. This hexasaccharide contains the H-type 3 antigen (Fucα1-2Galβ1-3GalNAc), previously shown to be recognized by BC2L-CN in cell binding assays and in glycan arrays (Figure 1C) [10,18]. We measured a K_d of 26 μM, which classified Globo H as the best ligand to date for BC2L-CN. The thermodynamic contributions indicated an enthalpy-driven increase of affinity for the larger ligands, but also strong entropic penalties associated with Globo H ligand.

2.4. Resolution of rBC2L-CN2 Structure in Complex with Oligosaccharide

In order to better understand the specificity of BC2L-CN towards the motif Fucα1-2Galβ1-3 and to obtain its atomic basis, we co-crystallized rBC2L-CN2 with H-type 1 tetrasaccharide and Globo H hexasaccharide. Clusters of plate crystals were obtained in a few days using a concentrated solution of

sodium citrate. Single plates were cryoprotected using 2.5 M sodium malonate pH 5 and diffracted to high resolution. Both structures were solved by molecular replacement. The structure of the H-type 1 complex belongs to space group R32:h (H32, a = b = 42.7 Å, c = 308.6 Å) with one monomer in the asymmetric unit, whereas the one of the H-type 3 complex belongs to the C2 space group (a = 74.4 Å, b = 42.9 Å, c = 102.6 Å and β = 96.0°) with one trimer in the asymmetric unit. Statistics on data and refinement are summarized in Table 2. No crystal has been obtained to date with Lewis Y as a ligand.

Table 2. Data collection and refinement statistics.

	H Type 1		H Type 3 (Globo H)	
	Data Collection			
Beamline	FIP-BM30A (ESRF)		Proxima 1 (Soleil)	
Wavelength	0.98096		0.97857	
Space group	H32/R32 (H)		C2	
Unit cell dimensions (Å,°)	a = b = 42.7, c = 308.6		a = 74.4, b = 42.9, c = 102.6, β = 96.0	
Resolution (Å)	36.71–1.61 (1.64–1.61)		37.10–1.90 (1.94–1.90)	
Nb/nb unique reflections	113,880/14,606		87,318/25,081	
R_{merge}	0.049 (0.588)		0.077 (0.363)	
R_{meas}	0.056 (0.643)		0.105 (0.486)	
Mean I/σI	23.9 (3.7)		8.7 (2.9)	
Completeness (%)	99.7 (97.5)		97.9 (97.0)	
Redundancy	7.8 (7.7)		3.5 (3.5)	
CC 1/2	0.999 (0.874)		0.995 (0.861)	
	Refinement			
Resolution (Å)	36.71–1.61		37.10–1.90	
Nb/nb free. reflections	14,605/751		25,080/1552	
R_{work}/R_{free}	15.8/20.3		16.5/22.7	
Rmsd Bond lengths (Å)	0.014		0.015	
Rmsd Bond angles (°)	1.78		2.0	
Rmsd Chiral ($Å^3$)	0.099		0.102	
No. atoms/Bfac ($Å^2$)	Chain A	Chain A	Chain B	Chain C
Protein	993/19.6	994/26.0	1027/25.9	1002/26.2
Ligand	36/23.2	47/33.6	58/41.6	58/38.8
Waters	135/28.3	109/32.2	106/31.7	82/32.5
Ramachandran Allowed (%)	100		100	
Favored (%)	97.8		96.9	
Outliers (%)	0		0	
PDB Code	6TID		6TIG	

Values in parentheses are for highest-resolution shell.

2.4.1. Overall Structure

The trimer formed in each complex (crystal symmetry were applied to build the trimer for the H-type 1 complex) was almost equivalent with a root mean square deviation (rmsd) of 0.16 Å for 393 aligned residues and was very similar to the trimer of the previous construct with an rsmd of 0.48 Å for 379 aligned residues [19]. Only minor differences were noticed on the bottom face of the jelly roll presenting the N- and C-termini of rBC2L-CN2. Both ends interacted through hydrogen bonds. Those interactions would not exist in the native BC2L-CN, since they involved residues from the added TEV cleavage site that remained after cleavage of the protein fusion. Nevertheless, they appeared to stabilize the domain. The small subsequent changes in the conformation of the termini, in particular for Trp127, impacted the nearby surface loops (Val28-Gly35, Asn64-Gln65, and Val96-Thr101) and resulted in small rigid body movements from 0.6 to 1.2 Å. Flipping of Gly97 was also observed, leading to alternative orientations of Ser98. In the H-type 1 complexed structure, the electron density for Ala31, Gly32, and Ser98 was too poor to allow modeling with suitable geometry. So, those residues were

omitted in the final coordinates (Figure 2). No changes relative to the original construct were observed on the top face of the jelly roll where the binding site was located.

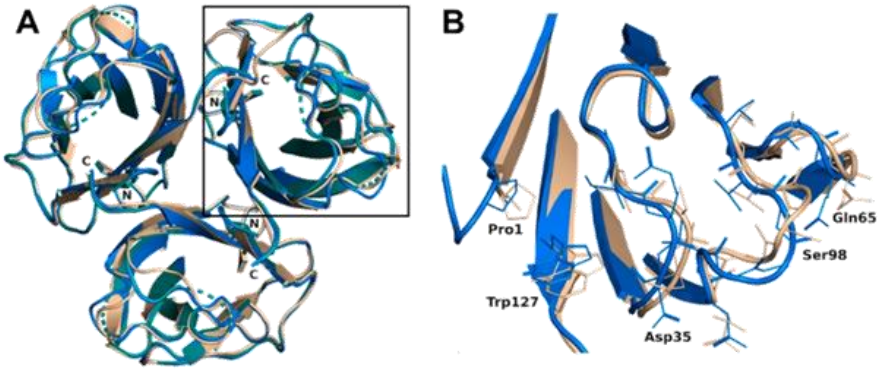


Figure 2. (**A**) Overlay of the overall trimer of rBC2L-CN2 in the complex with H-type 1 (green, 6TID, generated by crystal symmetry) or Globo H (blue, 6TIG) with the original trimer (beige, 2WQ4). (**B**) Zoom on the loop conformation changes observed in the protein chain B in the complex with Globo H.

2.4.2. Oligosaccharide Binding Interactions

Examination of the first electron density revealed without ambiguity binding of the oligosaccharides. Regarding the structure in complex with H-type 1 tetrasaccharide, only three of the four sugar units could be correctly modelled. The reducing galactose was too disordered as a result of its exposition to the solvent and the quality of the residual electron density was insufficient for modelling. For the structure in complex with Globo H hexasaccharide, the four units were modelled in every protein chain and the fifth one only in protein chains B and C. No electron density was visible for the reducing glucose, which, likewise, was completely solvent exposed (Figure 3).

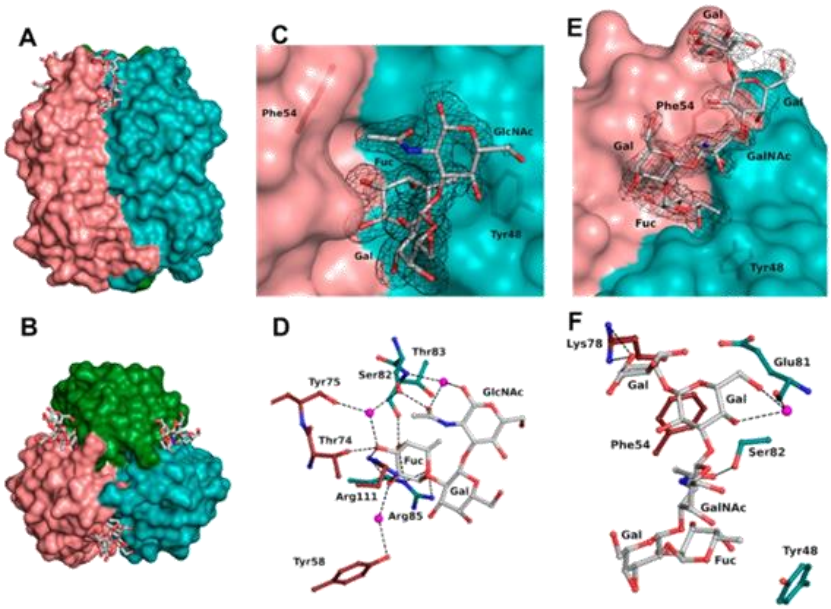


Figure 3. BC2L-CN2 binding site. Surface representation of the trimer of BC2L-CN2 in complex with Globo H, view from the side (**A**) and from the top (**B**). Surface representation of the binding interface between protomers B and C with the electron density displayed around H-type 1 (**C**) and Globo H (**E**) antigens at 1 σ (0.43 and 0.37 eA^3, respectively. Zoom on the interactions of BC2L-CN2 with H-type 1 (**D**) and Globo H (**F**) antigens. Protein residues are colored by protomer chain. Waters are represented as spheres.

The binding site was found at the interfaces between two protomers and was located on the top side of the jelly roll β-sandwich in a shallow groove (Figure 3A,B). The bottom of the binding site was principally made from residues from a surface loop (Glu81-Arg85) and also Tyr48. The orientation of the oligosaccharide was imposed by a wall made from Asn53, Phe54, Tyr58, Thr74, Lys78, Val110, and Arg111 from the neighboring protein chain. The fucose residue was buried in the protein, whereas the other sugar units of the oligosaccharide were pressed against the protein wall on one side while the other one was completely exposed to the solvent. As previously observed, the fucose was involved in extensive H-bonding with the protein and accounted for the majority of the interactions (Table 3 and Figure 3D). All oxygen atoms established direct or water-mediated contacts with the side chain of Arg85 (O4 and O5) and Ser82 (O3 through a water molecule) and the main chain of Thr83 (O4) from one protomer, as well as with the side chain of Arg111 (O2 and O3), Thr74 (O3), Tyr58 (O2 through a water molecule), and the main chain of Tyr75 (O3 through a water molecule) from the neighboring protomer involved in the binding site interface. Additionally, a hydrophobic interaction between the C6 methyl and the aromatic ring of Tyr48 as well as the two waters molecules that mediate hydrogen bonding were observed, matching what was observed in the previous crystal structure (2WQ4) [10].

Table 3. Summary of the interactions of BC2L-CN2 with oligosaccharides.

Ligand Atom	Protein Atom or Water	Distance (Å) H-Type 1	Distance (Å) H-Type 3
		Fuc1	
O2	Arg111 * NH2	3.06	3.04 ± 0.07
	Arg111 * NH1	2.88	3.03 ± 0.08
	HOH1 → Tyr 58 * OH	2.72 → 2.80	2.62 ± 0.08 → 3.01 ± 0.01
O3	Arg111 * NH2 *	3.18	3.15 ± 0.04
	Thr74 * OG1	2.54	2.62 ± 0.06
	HOH2 → Ser82 OG	2.59 → 2.66	2.60 ± 0.05 → 2.43 ± 0.03
	HOH2 → Tyr75 * O	2.59 → 2.64	2.60 ± 0.05 → 2.78 ± 0.01
O4	Arg85 NE	2.89	2.93 ± 0.02
	Thr83 O	2.69	2.71 ± 0.02
O5	Arg85 NH2	2.97	3.03 ± 0.08
C6	Tyr48	hydrophobic	hydrophobic
		GlcNAc3	
O7	Ser82 OG	2.57	2.78 ± 0.07
	HOH3 → Thr83 N	3.01 → 3.02	3.00 ± 0.33→ 3.03 ± 0.13 (not in chain C)
N-acetyl	Tyr54 *	hydrophobic	hydrophobic
		Gal4	
O4	HOH4 → Glu81 O		3.17 → 2.67 (chain C)
O6	HOH4 → Glu81 O		2.90 → 2.67 (chain C)
C1-C2	Phe54 *	hydrophobic	hydrophobic
		Gal5	
O6	Lys78 * NZ		2.90 (chain C)

Residues from the neighboring protomer in the binding interface are labelled with an asterisk (*). For water-mediated interaction, an arrow is used to indicate which water is linked to which protein atom. For Globo H complex, mean distance and standard deviations were calculated from the distance in each protomer. Only distances less than 3.2 Å are listed.

The next sugar residue, an α1-2 linked galactose (Gal2), was very solvent-exposed and did not interact with the protein. The following N-acetylated sugar, GlcNAc in H-type 1 or GalNAc in H-type 3, interacted mainly with its acetyl group. The acetyl oxygen was hydrogen-bonded to the side chain of Ser82 while the methyl group was engaged in a hydrophobic interaction with the aromatic ring of Phe54 from the neighboring protein chain. In some chains, the acetyl group's oxygen also established a hydrogen bond mediated by a water molecule with both the next glycosidic oxygen and the main

chain of Thr83. For the remaining sugar units in the Globo H structure, most of the interactions were seen in protein chain C and are depicted in Figure 3F. The O6 and O4 hydroxyls of the fourth sugar (Gal4) made a water-mediated hydrogen bond with the main chain oxygen of Glu81. In the α1-4 linked galactose (Gal5), only one interaction was observed between the O6 and the side chain of Lys78 from the neighboring chain. Those interactions with Gal4 and Gal5 were not seen in the protein chain B as a result of a different orientation of the O6 of Gal4 and of the terminal nitrogen of Lys78.

All sugar units presented the expected chair conformation and a good fit in the electron density as checked in Privateer [20]. The dihedral angles of glycosidic linkages were compatible with low-energy conformation, apart from the Galα1-4Gal connection that was in the vicinity of the second main low-energy minimum (Table 4) [21]. The Fucα1-2Gal linkage adopted dihedral angles consistent with the primary low-energy conformation, while for the other linkage, they corresponded to those of the secondary low-energy minimum.

Table 4. Dihedral angle measured for the observed glycosidic linkages of H-type 1 and Globo H.

	H-Type 1	Globo H		
Chain		**A**	**B**	**C**
	Fucα1-2Gal	Fucα1-2Gal		
Φ *	−78.9	−76.9	−78.8	−80.9
Ψ *	−101.5	−99.3	−100.5	−98.5
	Galβ1-3GlcNAc	Galβ1-3GalNAc		
Φ *	−73.3	−79.1	−66.9	−76.7
Ψ *	124.3	120.1	117.7	113.5
		GalNAcβ1-3Gal		
Φ *	NA	−78.8	−104.7	−92.6
Ψ *	NA	101.3	155.7	147.1
		Galα1-4Gal		
Φ *	NA	NA	72.2	79.1
Ψ *	NA	NA	98.1	100.5

* Φ (O5-C1-Ox-Cx) and Ψ (C1-Ox-Cx-Cx+1), in degrees (°) with x, the number of the carbon atoms of the second monosaccharide with which the 1 → x glycosidic bond was formed. NA: not applicable.

The overlay of all the binding sites shows that the conformation obtained for the H-type 1 trisaccharide was the one mainly observed in the H-type 3 trisaccharide (Figure 4). In protein chain C, there was a slight rotation around the C1-O2 bond of the first glycosidic linkage by ~22°, which led to a rigid body shift of the subsequent saccharide units Gal2, GalNAc3, and Gal4 by ~0.9 Å. This did not, however, impact strongly on ligand binding since both the fucose and the N-acetyl of GlcNAc or GalNAc that were involved in the major conserved interactions presented equivalent position. Only the aforementioned water-mediated interaction between the oxygen of the acetyl group and the glycosidic oxygen was lost. The GalNAc hydroxymethyl group had a different orientation in each protein chain and only one was involved in the water-mediated hydrogen bonds described above. The fifth sugar unit, Gal5, was well superposed in the two chains where it was observed, despite the rigid body shift observed for the other residues. In protein chain A no density was visible after Gal4, which presented different torsion angles than in the other chains, resulting in a more solvent-exposed sugar (Table 4).

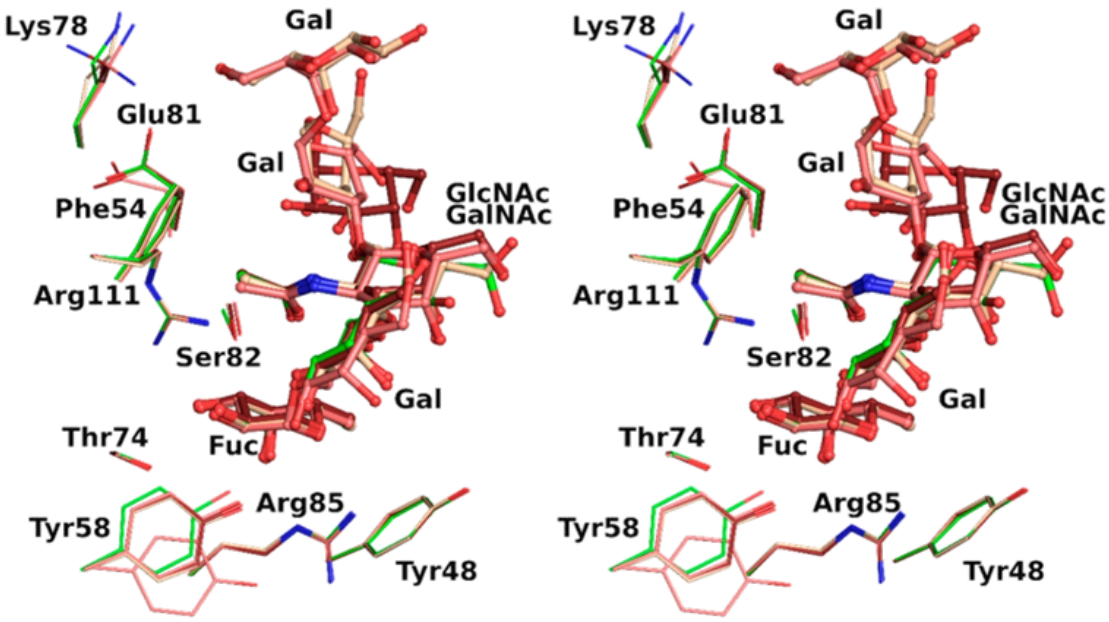


Figure 4. Overlay in wall eye stereo of each binding site observed in rBC2L-CN2 structure with H-type 1 (green) or Globo H (H-type 3) with carbon atoms colored in dark red, beige, and pink for chain A, B, and C, respectively.

3. Discussion

We designed a novel construct for BC2L-CN to allow crystallographic studies with oligosaccharides and to improve stability of the recombinant domain. Another construct encompassing only BC2L-CN residues created by Tateno et al. doesn't provide details on its length or the final protein yield [22]. Consequently, we chose to add a cleavable N-terminal fusion to help protein expression and allow purification by affinity on a nonsugar-based matrix. The rBC2L-CN2 met our expectations in terms of stability, as it is now stable over a year at 4 °C. The extra 50 amino acids of the original construct seem, therefore, to be indeed responsible for the precipitation formerly observed after merely two weeks of storage. The trimer was formed correctly and the overall structure was marginally influenced by the interactions between the new N- and C-termini. Conversely, these termini locally influenced the conformation of neighboring surface loops, but this did not reach or alter the binding site functionality, as confirmed by ITC measurements. The latest revealed a correlation between gain in affinity and oligosaccharide length that is not uncommon for bacterial lectins but remains to be fully clarified [23]. As is usually observed for lectin–carbohydrate pairs, the interaction was driven by enthalpy, with unfavorable entropy. The enthalpic gain can be linked to the novel interactions established by the additional carbohydrate moieties, while the entropic gain observed could be linked to the loss of flexibility upon binding or to desolvation effects. Nevertheless, we showed that BC2L-CN had the strongest affinity for Globo H, which was listed as one of the best ligands of BC2L-CN on glycan arrays, along with H-type 1 and Lewis Y [10].

Using rBC2L-CN2, we were able to obtain molecular information on the binding of two relevant oligosaccharides, to better characterize the protein carbohydrate binding site, where to date only the fucose subsite has been depicted [10]. The structures of rBC2L-CN2 in complex with H-type 1 and Globo H allowed us to identify a network of interactions composed of hydrogen bonds and hydrophobic contacts also implicating structural waters molecules, often involved themselves as mediators of protein–ligand interactions (Table 3). As observed in Figure 4, the full binding site was quite extended and solvent exposed and accommodated up to five carbohydrate units. The additional sugar moieties had little interaction with the protein surface but still allowed the identification of novel protein determinants and water molecules to be taken into consideration for ligand design, such as

Phe54, Lys78, and Glu81. It is striking to notice that with only one strong hydrogen bond and one hydrophobic contact (belonging to the N-acetyl group of GlcNAc/GalNAc), the affinity was improved by two orders of magnitude, from 2.7 mM for L-fucose to 55–77 μM for the tetrasaccharide. Equally striking is to note that no interactions were made by Gal2 and that the ones observed for Gal4 were not conserved: Rather than interacting with the residues of the binding site, these additional sugars seemed to play the role of a frame, allowing a handful of essential interactions to fall into place.

Both the H-type 1 and Globo H antigens are present on pluripotent cells and can be used as glyco-markers. The Globo H glycosphingolipid is also highly expressed on malignant tissues, particularly in breast and small cell lung cancers [24,25]. Globo H presents the H-type 3 antigen (Fucα1-2Galβ1-3GalNAc) which only differs from H-type 1 antigen (Fucα1-2Galβ 1-3GlcNAc) by the epimerization of GlcNAc at C4 position. Since no interaction implying the O4 hydroxyl is observed, the recognition of both antigens is equivalent at the structural level. Accordingly, BC2L-CN is an excellent probe for detecting or labelling undifferentiated human pluripotent stem cells or human embryonic stem cells [18,22,26]. Thus, the newly created BC2L-CN2 has great potential as a molecular tool for the detection of the Fucα1-2Galβ 1-3GlcNAc/GalNAc motif.

Unfortunately, no crystal complex could be obtained with the Lewis Y antigen (Fucα1-2Galβ1-4(Fucα1-3)GlcNAc), another high-affinity ligand of BC2L-CN that contains two terminal fucose moieties. Since H-type 2 was not well recognized by the protein, it is most probable that BC2L-CN bound preferentially the Fucα1-3 and not the Fucα1-2 presented by Lewis Y [10]. More insights on the binding mode of Lewis Y will come with further experimentation: Measurements of the affinity with the disaccharide Fucα1-3GlcNAc and continued crystallization trials.

The information gathered here can be used for the design of fucose-based glycodrugs towards BC2L-CN, since elimination of drug-resistant pathogens, such as *B. cenocepacia*, is still a critical issue nowadays, as available medication is limited. Numerous pathogens, in particular bacteria, exploit interactions between host-associated glycans and lectins for cell invasion and infection persistence by mediating adhesion to the host cells, and for pathogenesis by defining cell and tissue tropism [7,27]. Blocking the attachment of the pathogen to the host using lectin antagonists is a current route explored for the development of novel antimicrobial molecules (reviewed in [12,28,29]). Lectin inhibitors can also be used as antibiofilm molecules or for biofilm imaging [30,31]. Some lectins play indeed an important role in the formation and the maintenance of biofilms, a hallmark of chronic infections and resistance against antimicrobials (antibiotics or antifungals).

Following this theme, the structural information related to the spatial arrangement of the different binding sites (referred as lectin topology) can now facilitate the design of selective structure-based multivalent inhibitors. Multivalency, a common lectin feature, usually counterbalances the weak affinities observed in carbohydrate–lectin interactions, such as the values we have obtained for monosaccharides. Mediating several weak binding events simultaneously usually results in strong avidity and, therefore, increased affinity. Based on our structural data on BC2L-CN, a multivalent compound should present a fucose mimic to insure specificity; furthermore, the addition of an aglycone binding to the nonfucose subsites, in particular the one occupied by Gal5, could improve selectivity and affinity. Branching could be also attempted on the other side of the fucose binding site. As BC2L-CN is trivalent with the fucose binding sites 25 Å apart and at 11 Å from the top face of the trimer, a linker of appropriate length between the fucoside moieties and a scaffold for multivalency can now be rationally chosen.

As a result of its dual specificity and structural features, the superlectin BC2L-C from *B. cenocepacia* is indeed an interesting target for the design of antiadhesive or antibiofilm inhibitors. The dual carbohydrate specificity of this hexameric protein hints at its involvement in cross-linking. It presents several opposing binding surfaces: The N-terminal domain trimers (two, facing top and bottom of the hexamer) are selective for fucosides; conversely, the C-terminal domain dimers (three, forming a central belt) specifically bind to mannosides [11]. BC2L-C C-terminal domain would allow binding of the lectin to the bacterial cell wall through recognition of manno-configured carbohydrates, as

observed for its homolog BC2L-A, another soluble lectin from *B. cenocepacia* [32]. On the other hand, the N-terminal domain would target fucosylated ligands on host cells, in particular human blood group oligosaccharides. BC2L-A and especially LecB have been targets for the design of both mono- and multivalent antiadhesive compounds for over a decade. They are derived either from fucose or mannose (reviewed recently in [12]) and could be effective on BC2L-C. Nevertheless, to date only C-fucosides-calix[4]arene (1-3-alternate) have been tested on BC2L-C [33]. This inhibitor is able to crosslink *B. cenocepacia* cells and inhibits BC2L-C-induced hemagglutination. As the compound does not inhibit hemagglutination induced by LecB, it most probably binds to BC2L-CN, suggesting also a role of this domain into cell cross-linking and, hence, in biofilm formation. Since both BC2L-C domains have millimolar affinity for L-fucose (personal communication of O. Sulak), more studies seem necessary to understand the mode of binding of inhibitors and to investigate the role of both domains in the biofilm of *B. cenocepacia*.

Key molecular determinants involved in the binding of BC2L-CN to human oligosaccharides are now identified. Further probing will permit us to complete our understanding of BC2L-CN's binding to human epitopes and, in particular, to Lewis Y. The data gathered here are being used for the conception of glycodrugs specific to BC2L-CN, which will then be synthetized, and their potential evaluated as antiadhesives or antibiofilm agents.

4. Materials and Methods

4.1. Protein Expression and Purification

The DNA sequence encoding for BC2L-C-Nter comprising amino acids 2 to 132 was amplified by PCR with purposely designed primers using previous construct as template and 5′- CTTCATATGCCGCTGCTGAGCGCCAGTATCG-3′ and 5′-TACTCGAGTTATGCCGCGGTGCC CCAAATCG-3′ as forward and reverse primers, respectively (restriction sites are underlined) [10]. The PCR product (ca. 400 base pairs) was purified from 1% agarose gel using Nucleospin Gel and PCR Clean-up kit (Macherey-Nagel, Hoerdt, France) using manufacturer instructions. The gene product and homelab vectors of interest (pET-TEV [14] and pCold-TEV [16]) were digested with NcoI and XhoI restriction enzymes (New England Biolabs, Evry, France) for 1 h at 37 °C, purified, and ligated at room temperature using the Takara mix. The pCold-TEV originates from the pCold-TF vector (Takara Bio Europe, Saint Germain en Laye, France) where the enterokinase site was replaced by tobacco etch virus (TEV) cleavage site by PCR using the 5′-cgcggtagtggtggtgaaaacctgtattttcagggccatatggagctcggtacc-3′ and 5′-accaccactaccgcgtggcaccagacccgc-3′ as forward and reverse primers, respectively, with the PrimeSTAR Max DNA polymerase (Takara Bio Europe, Saint Germain en Laye, France) according to manufacturer instructions.

After transformation by heat shock, *Escherichia coli* BL21 Star (DE3) cells harboring the plasmid were cultured in Luria Broth (LB) broth medium supplemented with 100 µg/mL ampicillin at 37 °C under constant shaking. At OD_{600nm} = 0.4, the incubator temperature was decreased to 16 °C and when OD_{600nm} reached 0.7, the protein expression was induced overnight by the addition of 0.1 mM IPTG. Then, cells were centrifuged at room temperature, 5 min at 5000× *g*, and the resulting pellet was weighed. Each g of wet cell pellet was resuspended with 5 mL of Buffer 1 (Tris-HCl 50 mM, NaCl 100 mM, pH 8.5) prior to treatment with DENARASE® endonuclease (c-LEcta GMBH, Leipzig, Germany) for 10 min at room temperature on a rotating wheel. The cells were lysed by pressure at 1.9 MPa using a one-shot table-top cell disruptor (Constant Systems Ltd.). The lysate was centrifuged 30 min, 24,000× *g* at 4 °C, and the resulting supernatant filtered through a 0.45 µm polyethersulfone (PES) syringe filter prior to loading on a 5 mL HisTrap™ fast flow (FF) column (GE Heathcare Life Sciences, Marlborough, MA, USA) equilibrated with buffer 1 for affinity chromatography using NGC system (Bio-Rad, Marnes-la-Coquette, France)). After washing the unbound proteins with buffer 1, rBC2L-CN2 was eluted using a 20 column volumes (CV) gradient of 0–500 mM imidazole. Fractions containing the protein were pooled after examination on 15% SDS-PAGE gel. The imidazole was

removed using a PD10 desalting column (GE Healthcare Life Sciences, Marlborough, MA, USA). The protein was concentrated by centrifugation (Vivaspin 3KDa, Sartorius, Goettingen, Germany) to at least 0.7 mg.mL^{-1} before addition of TEV protease (1:50 *w/w*, enzyme:protein ratio), 1 mM ethylenediaminetetraacetic acid (EDTA) and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) for tag cleavage overnight at 19 °C [15]. The sample was again submitted to affinity chromatography (same conditions as previously) to separate two fragments of 14 kDa and 52 kDa corresponding to the target protein and its cleave fusion, respectively (assessed by SDS–PAGE). After concentration by centrifugation as previously described, the protein concentration was determined by UV absorbance at 280 nm with a NanoDrop 2000 spectrophotometer (Thermo Scientific, Illkirch-Graffenstaden, France).

SEC was performed on an ENrichTM SEC 70 10 × 300 column (Bio-Rad, Marnes-la-Coquette, France)) using a NGC™ systems (Bio-Rad Ltd.). The analytical column was pre-equilibrated with 20 mM Tris-HCl pH 7.0 and 100 mM NaCl, optimized for protein stability via thermal shift assay (TSA). The volume for the sample injections was 240 µL and the flow rate was 1.0 mL/min. A column calibration curve using gel-filtration standards (GE Healthcare, Life Sciences) was performed to allow the calculation of the protein molecular weight.

4.2. ITC Measurements

All experiments were performed at 25 °C with an ITC200 isothermal titration calorimeter (Microcal-Malvern Panalytical, Orsay, France). The rBC2L-CN2 and sugars were dissolved in the same buffer composed of 100 mM Tris HCl pH 7.0 and 100 mM NaCl. A total of 20 to 38 injections of 1 µL of sugar solution (10 or 15 mM) were added at intervals of 100 or 200 s while stirring at 850 rev/min^{-1} in the 200 µL sample cell containing the protein, at 250 or 340 µM. The experimental data were fitted to a theoretical titration curve using the supplied software Origin 7. They permitted us to determine in one experiment affinity (i.e., association constant, Ka), binding enthalpy (ΔH), and stoichiometry (n). Free energy change (ΔG) and entropy contributions ($T\Delta S$) were derived from the equation $\Delta G = \Delta H - T\Delta S = -RT \ln Ka$ (with T the absolute temperature and R = 8.314 J mol^{-1} K^{-1}. For experiments in ligand excess, the stoichiometry was fixed to 1. Two experiments were performed for L-galactose, three for Globo H hexasaccharide (H-type 3), and only one for H-type 1 tetrasaccharide or Lewis Y pentasaccharide. The oligosaccharides were purchased from Elicityl, Crolles, France.

4.3. Crystallization, Data Collection, and Structure Determination

Oligosaccharides (H-type 1 tetrasaccharide and Globo H hexasaccharide) at 10 mM in water were added to rBC2L-CN2 at a concentration of 5 mg.mL^{-1} such that the final ligand concentration was 1 mM. After incubation at room temperature (22 °C) for at least 1h, crystallization conditions were screened using the vapor diffusion method and 2-µL hanging drops containing a 50:50 (*v/v*) mix of protein and reservoir solution. The screens used included: BCS Eco Screen, Eco Structure Screen 2, Morpheus I-carboxylic acids, and MIDAS (Molecular Dimensions Ltd., Sheffield, UK). Crystals were obtained in a few days from solution 48 of the Structure Screen 2 and optimized using 1.2–1.4 M sodium citrate pH 7.0. Both complexes led to clusters of plates which were broken, transferred to 2.5 M sodium malonate (CryoProtX, Molecular Dimensions Ltd.) for cryoprotection, and flash-cooled in liquid nitrogen prior to data collection. H-type 1 complexed data were collected at European Synchrotron Radiation Facility (ESRF), Grenoble France, on beamline FIP-BM30A using a ADSC Q315r detector (Area Detector Systems Corporation, Poway, CA, USA),while those for the Globo H complex were collected on the beamline Proxima 1, synchrotron SOLEIL, Saint Aubin, France, using an Eiger 16 m detector (Dectris, Baden, Switzerland).

The data were processed using XDS and XDSME [34,35]. All further computing was performed using the CCP4 suite [36]. The coordinates of the monomer of PDB-ID 2WQ4 were used as a search model to solve both complexed structures of rBC2L-CN by molecular replacement using PHASER [37]. Refinement was performed using restrained maximum likelihood refinement and REFMAC 5.8 [38] iterated with manual rebuilding in Coot [39]. Five percent of the observations were set aside for

cross-validation analysis with the same set used for both structures. Hydrogen atoms were added in their riding positions and used for geometry and structure-factor calculations. The final model was validated with the wwPDB Validation server, https://validate-rcsb-1.wwpdb.org/, and the sugar conformation was checked in Privateer [20]. The coordinates were deposited in the Protein Data Bank (PDB) under codes 6TID and 6TIG for H-type 1 and Globo H complex structures, respectively. Data and refinement quality statistics are summarized in Table 1.

Author Contributions: Conceptualization, A.V.; methodology, A.V. and R.B.; formal analysis, A.V. and R.B.; writing—original draft preparation, R.B.; writing—review and editing, A.V., A.B., and R.B.; visualization, A.V., A.B., and R.B.; supervision, A.V. and A.B.; project administration, A.V.; funding acquisition, A.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 765581 and the French national research agency (ANR) projects Glyco@Alps (ANR-15-IDEX-02).

Acknowledgments: Thanks to Valérie Chazalet for preparing the pCold-TEV vector and to Emilie Gillon for her help in ITC. The authors are grateful for access to the beamlines FIP-BM30A at the European Synchrotron Radiation Facility, Grenoble, France, (proposal number 20180234/20180234) and Proxima 1 at SOLEIL Synchrotron, Saint Aubin, France, (proposal number 20170827). We also would like to thank our local contacts Jean-Luc Ferrer, Serena Sirigu, and Pierre Legrand for their assistance and technical support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Fazli, M.; Almblad, H.; Rybtke, M.L.; Givskov, M.; Eberl, L.; Tolker-Nielsen, T. Regulation of biofilm formation in *Pseudomonas* and *Burkholderia* species. *Environ. Microbiol.* **2014**, *16*, 1961–1981. [CrossRef] [PubMed]
2. Rhodes, K.A.; Schweizer, H.P. Antibiotic resistance in *Burkholderia* species. *Drug Resist Updat.* **2016**, *28*, 82–90. [CrossRef] [PubMed]
3. Parajuli, N.P.; Acharya, S.P.; Mishra, S.K.; Parajuli, K.; Rijal, B.P.; Pokhrel, B.M. High burden of antimicrobial resistance among gram negative bacteria causing healthcare associated infections in a critical care unit of Nepal. *Antimicrob. Resist. Infect. Control* **2017**, *6*, 67. [CrossRef] [PubMed]
4. Scoffone, V.C.; Chiarelli, L.R.; Trespidi, G.; Mentasti, M.; Riccardi, G.; Buroni, S. *Burkholderia cenocepacia* Infections in Cystic Fibrosis Patients: Drug Resistance and Therapeutic Approaches. *Front. Microbiol.* **2017**, *8*, 1592. [CrossRef]
5. Mahenthiralingam, E.; Urban, T.A.; Goldberg, J.B. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat. Rev. Microbiol.* **2005**, *3*, 144–156. [CrossRef]
6. Sousa, S.A.; Ramos, C.G.; Leitão, J.H. *Burkholderia cepacia* Complex: Emerging Multihost Pathogens Equipped with a Wide Range of Virulence Factors and Determinants. *Int. J. Microbiol.* **2011**, *2011*, 607575. [CrossRef]
7. Imberty, A.; Varrot, A. Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.* **2008**, *18*, 567–576. [CrossRef]
8. Lameignere, E.; Malinovska, L.; Slavikova, M.; Duchaud, E.; Mitchell, E.P.; Varrot, A.; Sedo, O.; Imberty, A.; Wimmerova, M. Structural basis for mannose recognition by a lectin from opportunistic bacteria *Burkholderia cenocepacia*. *Biochem. J.* **2008**, *411*, 307–318. [CrossRef]
9. Heggelund, J.E.; Varrot, A.; Imberty, A.; Krengel, U. Histo-blood group antigens as mediators of infections. *Curr. Opin. Struct. Biol.* **2017**, *44*, 190–200. [CrossRef]
10. Sulák, O.; Cioci, G.; Delia, M.; Lahmann, M.; Varrot, A.; Imberty, A.; Wimmerová, M. A TNF-like Trimeric Lectin Domain from *Burkholderia cenocepacia* with Specificity for Fucosylated Human Histo-Blood Group Antigens. *Structure* **2010**, *18*, 59–72. [CrossRef]
11. Sulák, O.; Cioci, G.; Lameignère, E.; Balloy, V.; Round, A.; Gutsche, I.; Malinovská, L.; Chignard, M.; Kosma, P.; Aubert, D.F.; et al. *Burkholderia cenocepacia* BC2L-C is a super lectin with dual specificity and proinflammatory activity. *PLoS Pathog.* **2011**, *7*, e1002238. [CrossRef] [PubMed]
12. Meiers, J.; Siebs, E.; Zahorska, E.; Titz, A. Lectin antagonists in infection, immunity, and inflammation. *Curr. Opin. Chem. Biol.* **2019**, *53*, 51–67. [CrossRef] [PubMed]

13. Tamburrini, A.; Colombo, C.; Bernardi, A. Design and synthesis of glycomimetics: Recent advances. *Med. Res. Rev.* **2019**. [CrossRef] [PubMed]
14. Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R.W.; Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: Comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.* **2007**, *81*, 6807–6816. [CrossRef]
15. Kapust, R.B.; Tözsér, J.; Fox, J.D.; Anderson, D.E.; Cherry, S.; Copeland, T.D.; Waugh, D.S. Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng. Des. Sel.* **2001**, *14*, 993–1000. [CrossRef]
16. Qing, G.; Ma, L.-C.; Khorchid, A.; Swapna, G.V.T.; Mal, T.K.; Takayama, M.M.; Xia, B.; Phadtare, S.; Ke, H.; Acton, T.; et al. Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat. Biotechnol.* **2004**, *22*, 877–882. [CrossRef]
17. Bleuler-Martinez, S.; Stutz, K.; Sieber, R.; Collot, M.; Mallet, J.M.; Hengartner, M.; Schubert, M.; Varrot, A.; Kunzler, M. Dimerization of the fungal defense lectin CCL2 is essential for its toxicity against nematodes. *Glycobiology* **2017**, *27*, 486–500. [CrossRef]
18. Tateno, H.; Matsushima, A.; Hiemori, K.; Onuma, Y.; Ito, Y.; Hasehira, K.; Nishimura, K.; Ohtaka, M.; Takayasu, S.; Nakanishi, M.; et al. Podocalyxin is a glycoprotein ligand of the human pluripotent stem cell-specific probe rBC2LCN. *Stem Cells Transl. Med.* **2013**, *2*, 265–273. [CrossRef]
19. Krissinel, E.; Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2256–2268. [CrossRef]
20. Agirre, J.; Iglesias-Fernandez, J.; Rovira, C.; Davies, G.J.; Wilson, K.S.; Cowtan, K.D. Privateer: Software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **2015**, *22*, 833–834. [CrossRef]
21. Pérez, S.; Sarkar, A.; Rivet, A.; Breton, C.; Imberty, A. Glyco3D: A Portal for Structural Glycosciences. In *Glycoinformatics*; Lütteke, T., Frank, M., Eds.; Springer: New York, NY, USA, 2015; pp. 241–258. [CrossRef]
22. Tateno, H.; Toyota, M.; Saito, S.; Onuma, Y.; Ito, Y.; Hiemori, K.; Fukumura, M.; Matsushima, A.; Nakanishi, M.; Ohnuma, K.; et al. Glycome Diagnosis of Human Induced Pluripotent Stem Cells Using Lectin Microarray. *J. Biol. Chem.* **2011**, *286*, 20345–20353. [CrossRef] [PubMed]
23. Navarra, G.; Zihlmann, P.; Jakob, R.P.; Stangier, K.; Preston, R.C.; Rabbani, S.; Smiesko, M.; Wagner, B.; Maier, T.; Ernst, B. Carbohydrate–Lectin Interactions: An Unexpected Contribution to Affinity. *ChemBioChem* **2017**, *18*, 539–544. [CrossRef] [PubMed]
24. Breimer, M.E.; Säljö, K.; Barone, A.; Teneberg, S. Glycosphingolipids of human embryonic stem cells. *Glycoconj. J.* **2017**, *34*, 713–723. [CrossRef] [PubMed]
25. Lai, T.-Y.; Chen, I.J.; Lin, R.-J.; Liao, G.-S.; Yeo, H.-L.; Ho, C.-L.; Wu, J.-C.; Chang, N.-C.; Lee, A.C.-L.; Yu, A.L. Fucosyltransferase 1 and 2 play pivotal roles in breast cancer cells. *Cell Death Discov.* **2019**, *5*, 74. [CrossRef] [PubMed]
26. Onuma, Y.; Tateno, H.; Hirabayashi, J.; Ito, Y.; Asashima, M. rBC2LCN, a new probe for live cell imaging of human pluripotent stem cells. *Biochem. Biophys. Res. Commun.* **2013**, *431*, 524–529. [CrossRef] [PubMed]
27. Poole, J.; Day, C.J.; von Itzstein, M.; Paton, J.C.; Jennings, M.P. Glycointeractions in bacterial pathogenesis. *Nat. Rev. Microbiol.* **2018**, *16*, 440–452. [CrossRef] [PubMed]
28. Cecioni, S.; Imberty, A.; Vidal, S. Glycomimetics versus multivalent glycoconjugates for the design of high affinity lectin ligands. *Chem. Rev.* **2015**, *115*, 525–561. [CrossRef]
29. Sattin, S.; Bernardi, A. Glycoconjugates and Glycomimetics as Microbial Anti-Adhesives. *Trends Biotechnol.* **2016**, *34*, 483–495. [CrossRef]
30. Johansson, E.M.V.; Crusz, S.A.; Kolomiets, E.; Buts, L.; Kadam, R.U.; Cacciarini, M.; Bartels, K.-M.; Diggle, S.P.; Cámara, M.; Williams, P.; et al. Inhibition and Dispersion of Pseudomonas aeruginosa Biofilms by Glycopeptide Dendrimers Targeting the Fucose-Specific Lectin LecB. *Chem. Biol.* **2008**, *15*, 1249–1257. [CrossRef]
31. Wagner, S.; Hauck, D.; Hoffmann, M.; Sommer, R.; Joachim, I.; Müller, R.; Imberty, A.; Varrot, A.; Titz, A. Covalent Lectin Inhibition and Application in Bacterial Biofilm Imaging. *Angew Chem. Int. Ed.* **2017**, *56*, 16559–16564. [CrossRef]
32. Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; de Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; et al. *Burkholderia cenocepacia* lectin A binding to heptoses from the bacterial lipopolysaccharide. *Glycobiology* **2012**, *22*, 1387–1398. [CrossRef]

33. Kašáková, M.; Malinovská, L.; Klejch, T.; Hlaváčková, M.; Dvořáková, H.; Fujdiarová, E.; Rottnerová, Z.; Maťátková, O.; Lhoták, P.; Wimmerová, M.; et al. Selectivity of original C-hexopyranosyl calix[4]arene conjugates towards lectins of different origin. *Carbohydr. Res.* **2018**, *469*, 60–72. [CrossRef]
34. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132. [CrossRef]
35. Legrand, P. XDSME: XDS Made Easier. *GitHub Repos.* **2017**. [CrossRef]
36. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242. [CrossRef]
37. McCoy, A.J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 32–41. [CrossRef] [PubMed]
38. Murshudov, G.N.; Skubak, P.; Lebedev, A.A.; Pannu, N.S.; Steiner, R.A.; Nicholls, R.A.; Winn, M.D.; Long, F.; Vagin, A.A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367. [CrossRef] [PubMed]
39. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501. [CrossRef] [PubMed]

Sample Availability: Samples of rBC2L-CN2 in pCold-TEV are available from the authors.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

4.3. Further information

a. Structural Study

To follow up on the previous disclosure of novel BC2L-C-N_{ter}/oligosaccharide complexes, a new co-crystallized structure was obtained with Lewis y (Le^y, pentasaccharide) as ligand. Co-crystallization was achieved by the hanging-drop method using 1.2 M sodium citrate at pH 7.0, which generated clusters of crystal plates within 48h. The crystals were cryo-protected using 2.5 M sodium malonate at pH 5.0 and diffracted to high resolution. Molecular replacement was used to solve the structure at 1.92 Å. The statistics of data collection are presented in **Table 4.1**.

The affinity of BC2L-C-N_{ter} for Le^y is among the strongest measured to date (K_D = 52.6 µM). The study of a complex featuring Le^y gains relevance from the fact that this pentasaccharide, unlike other ligands of BC2L-C-Nt, presents a branching fucose residue: Fucα1-2Galβ1-4(**Fucα1-3**)GlcNAcβ1-3Gal. It was, thus, pertinent to study its binding mode, which could either be similar to what has been observed or involve the branched fucose residue.

Table 4.1. Data collection and refinement statistics. Values in parentheses are for highest-resolution shell.

Data Collection	
Beamline	Proxima 2A (Soleil)
Wavelength	0.98011
Space group	H32/R32 (H)
Unit cell dimensions (Å,°)	a = b = 42.9, c = 310.0
Resolution (Å)	36.88-1.92 (1.97-1.92)
Nb/nb unique reflections	139,837/8,958
R_{merge}	0.073 (0.399)
R_{meas}	0.077 (0.419)
Mean I/σI	23.8 (7.3)
Completeness (%)	100.0 (100.0)
Redundancy	15.6 (16.5)
CC 1/2	0.999 (0.982)
Refinement (in progress)	
Resolution (Å)	36.91-1.92
Nb/nb free. reflections	8957/429
R_{work}/R_{free}	14.9/19.2
Rmsd Bond lengths (Å)	0.019
Rmsd Bond angles (°)	1.96
Rmsd Chiral (Å^3)	0.481
No. atoms/Bfac (Å^2):	
Protein	987/28.7
Ligand	57/41.5
Waters	101/34.3
Ramachandran Allowed (%)	100
Favored (%)	96.2
Outliers (%)	0

Inspection of the protein/ligand interactions, detailed in **Figure 4.1** and **Table 4.2**, confirmed the known binding mode in prior crystal structures: the first unit **Fuc1** established H-bonds with residues **Thr74**, **Thr83**, **Arg85** and **Arg111**, as well as water-mediated contacts with **Tyr75**, **Ser82**, and **Tyr58** (crystallographic waters **1** and **2**, respectively) and a hydrophobic interaction between the C6 methyl group and the aromatic ring of **Tyr48**. While solving the structure, the possibility of an alternative binding mode featuring the branched unit **Fuc4** as principal interacting moiety was considered, to no avail. This proved that all oligosaccharide ligands studied to date bind to BC2L-C-N_{ter} through their terminal fucose. Nevertheless, new structural features exclusive to this crystal structure provided us with further insight for our study.

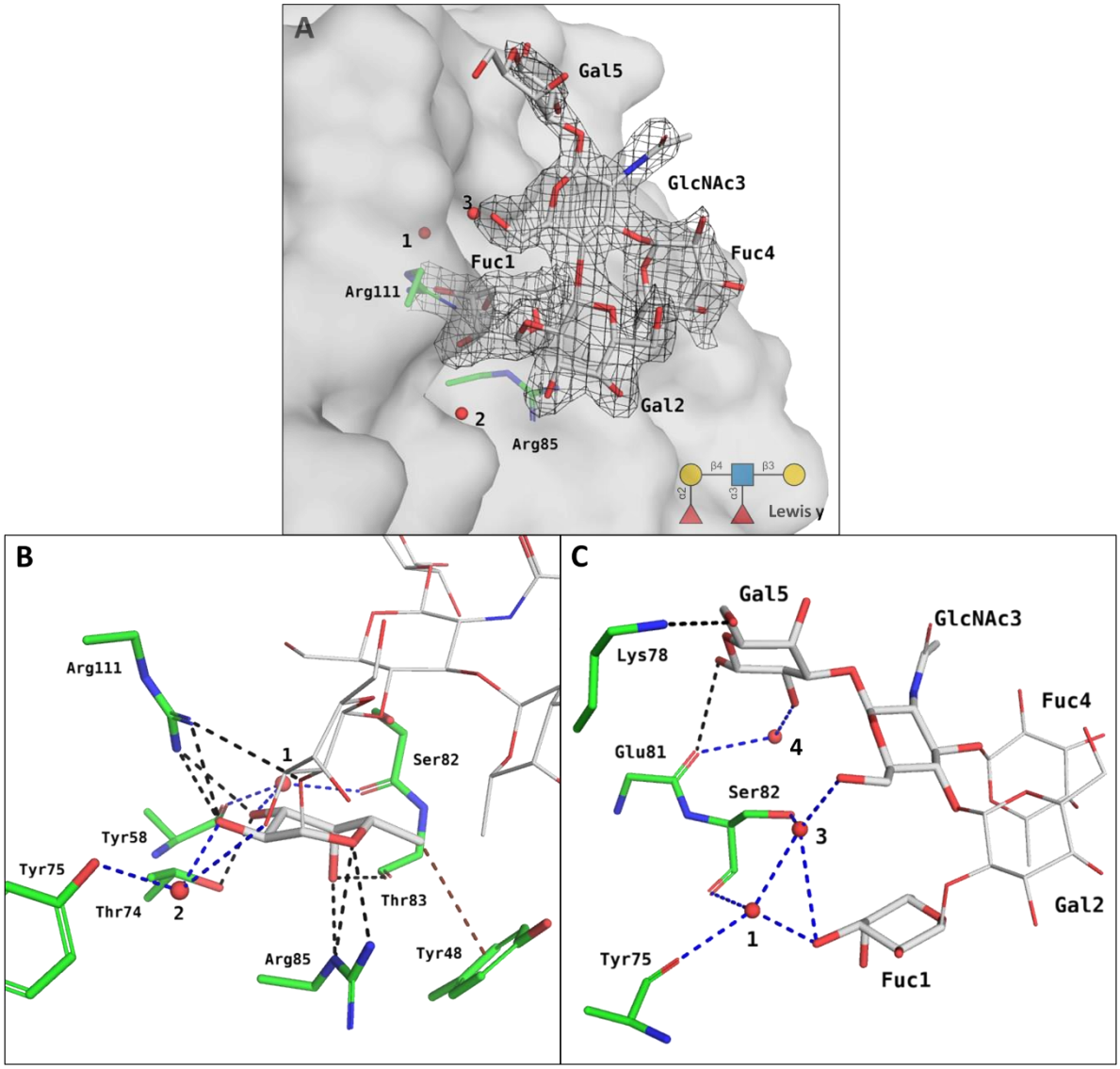


Figure 4.1. Crystal structure of the BC2L-C-Nt/Le^y complex. **A:** Overview of the ligand electron density for Le^y and its binding mode. **B:** Focus on the binding mode of unit **Fuc1**. **C:** New structural information provided by Le^y. Water molecules are depicted as red spheres, protein surface in transparent gray, protein/ligand interactions are depicted in black dashes, water-mediated contacts in blue, hydrophobic interaction in brown.

Table 4.2. Summary of the interactions observed between BC2L-C-N_{ter} and Lewis y.*

Ligand Atom	Protein Atom or Water	Distance (Å)
	Fuc1	
O2	Arg111 * NH2	3.06
	Arg111 * NH1	2.87
	HOH2 → Tyr 58 * OH	2.67 → 3.14
O3	Arg111 * NH2	3.04
	Thr74 * OG1	2.62
	HOH1 → Tyr75 * O	2.48 → 2.82
	HOH1 → HOH3	2.48 → 2.96
	(HOH1 → Ser82 O)	(2.48 → 3.41)
	(HOH1 → Ser82 OG)	(2.48 → 3.78)
	(HOH3)	(3.54)
O4	Arg85 NE	2.93
	Thr83 O	2.76
O5	Arg85 NH2	3.02
C6	Tyr48	hydrophobic
	GlcNAc3	
O6	HOH3 → HOH1	2.63 → 2.96
	HOH3 → Ser82 OG	2.63 → 2.65
	Gal5	
O6	Lys78 * NZ	2.98
O2	HOH4 → Glu81 O	2.78 → 2.89
O1	(Glu81) O	3.39
C5	Phe54 *	hydrophobic

*Residues from the neighbouring protomer in the binding interface are labelled with an asterisk (*). For water-mediated interactions, an arrow indicates which water is linked to which protein atom. Distances over 3.2 Å are listed in parenthesis.

As for H-type 1 and Globo H, the second unit **Gal2** was solvent-exposed and acted as a spacer or directing moiety for the third unit, without any direct interaction with the protein. The third unit **GlcNAc3** was mostly exposed to the solvent, except for the substituent that pointed towards the protein. In previous crystal structures, this substituent was the N-acetyl moiety (N-Ac), which established two contacts with the protein wall: its methyl group pointed towards the phenyl ring of **Phe54**, forming a hydrophobic interaction and its carbonyl oxygen H-bonded with **Ser82**'s hydroxyl group (see **Figure 4.2**). Compared with H-type 1 and Globo H, Le^y presents a different linkage between units 2 and 3: Galβ1-4GlcNAc instead of β1-3. Adding this to the presence of the branching **Fuc4** made clear that the N-Ac group could not

occupy the position previously described. Interestingly, H-bonding to **Ser82** was still observed, albeit different: water molecule **3** mediated an interaction between **Ser82**'s hydroxyl and **GlcNAc3**'s hydroxymethyl, which spatially replaced the acetyl group. Water molecule **3** can be considered a crystallographic water, as it is tightly flanked by protein and ligand, and coordinates **1**, another crystallographic water essential for the interaction. Interestingly, the side chain of **Ser82** shifts to accommodate this binding mode, resulting in a new setting for the usually well-conserved environment of water **1** (see distances in **Table 4.2**).

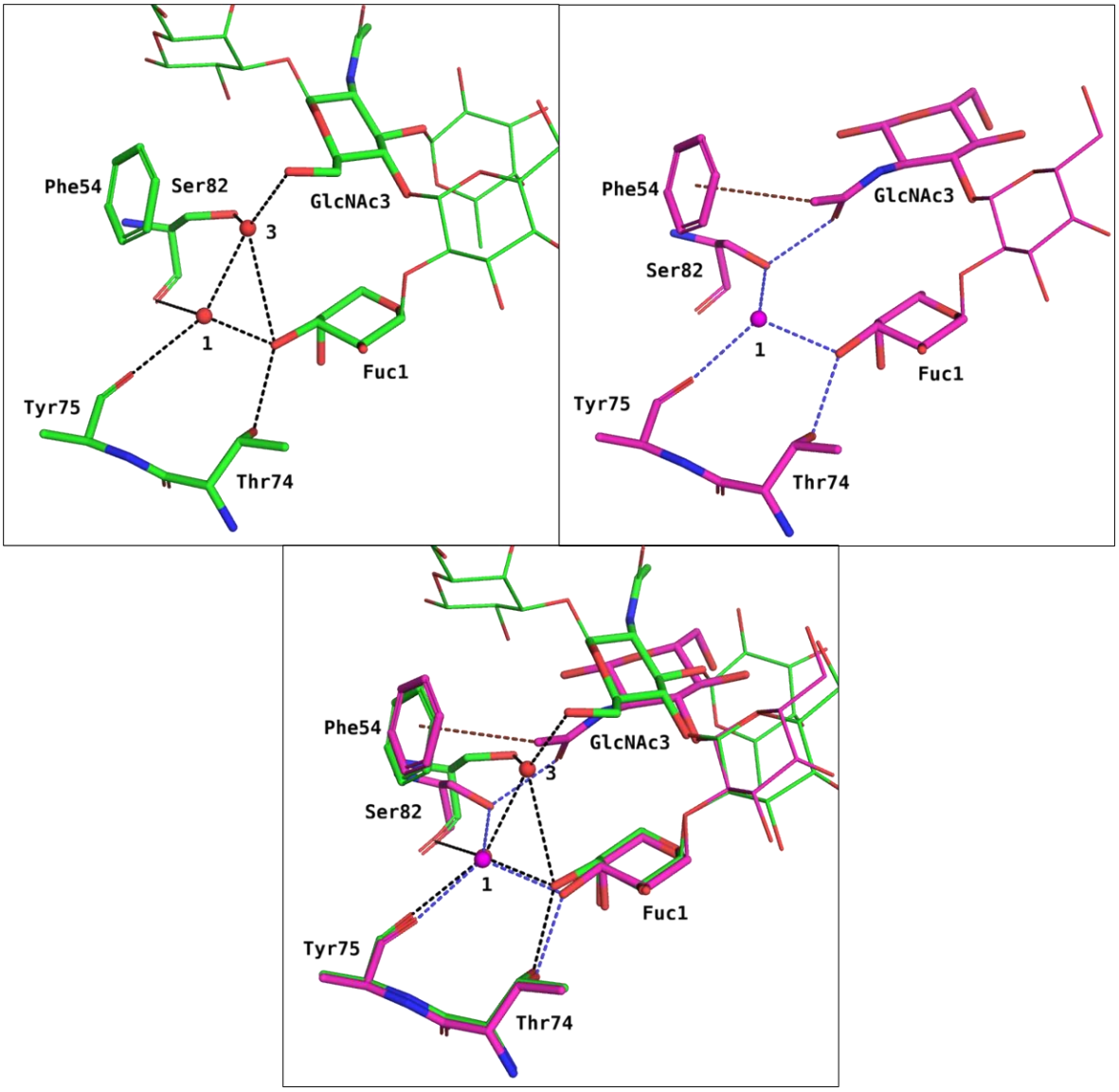


Figure 4.2. Comparison of binding modes: Lewis y (**top left**), H-type 1 (**top right**), and superposition (**bottom**). Water molecules are depicted in red or pink and interactions are depicted in black and blue for Lewis y and H-type 1, respectively. Hydrophobic interaction in brown.

Commonly observed in carbohydrate/lectin interactions, water 'sandwiched' between ligand and protein can result in tightly knit H-bond networks. Such surface level H-bonding can contribute up to -6 kJ/mol to the binding free energy ΔG, accounting for a 10-fold affinity

increase.[166] This would rationalize the affinity increase observed from monosaccharide to oligosaccharide ligands. Even though water **3** and its network aren't present in structures featuring N-Ac as the partner of **Ser82**, other networks featuring heavily coordinated waters were present (see section **4.2**, **Figure 3**). On the other hand, thermodynamic study (described below) shows micromolar affinity only upon adding the third sugar unit, without significant increase for the tetra-, penta-, or hexasaccharide. With this in mind, attributing the affinity increase to the hydroxymethyl/acetyl group's contacts to key residue **Ser82** is plausible.

Collectively, these observations allowed the rationalization the micromolar affinity of BC2L-C-N_{ter} for oligosaccharides bearing the scaffold Fucα1-2Galβ1-3*(/4)*Glc*(/Gal)*NAc. These ligands consistently establish the known contacts of the first sugar unit, while creating new water-mediated H-bond networks. The third unit (GlcNAc or GalNac) is likely responsible for the affinity jump from millimolar to micromolar by interacting with residue **Ser82** through either an N-Ac or CH_2OH moiety.

An additional crystal structure of the *apo*-form of the lectin was obtained and solved by Kanhaya Lal from the PhD4GlycoDrug consortium.[167] Analysis of this structure confirmed that the CRD of the lectin is structurally constant with or without ligand, and consistently houses crystallographic water molecules **1** and **2**.

b. Biophysical Evaluation

In order to complete the understanding of the target's affinity for oligosaccharide ligands, ITC was used to measure affinities for αMeFuc, Fucα(1-2)Gal, and Fucα(1-2)Galβ(1-3)GlcNAc (H-type 1 trisaccharide). Summarized in **Table 4.3**, this study conclusively rationalized the affinity jump observed from mono- to oligosaccharide ligands. The H-type 1 trisaccharide presents the strongest affinity for BC2L-C-N_{ter} to date (25 µM), which points at the third sugar unit **GlcNAc** as responsible for the 10-fold affinity increase from the disaccharide (2.5 mM). Thermodynamically, it is established that enthalpic gains are the main factor for this change, likely due to the additional contacts provided by GlcNAc. It terms of entropy, there doesn't seem to be a trend influencing affinity as the oligosaccharide chain grows. This contrasts with the entropy-related effects that have been observed for similar cases in which long oligosaccharide chains bind to lectins.[168]

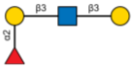
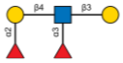
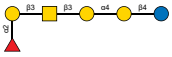
Ligand		N	K_D (μM)	−ΔG (kJ/mol)	−ΔH (kJ/mol)	−TΔS (kJ/mol)	Ref.
αMeFuc monosaccharide		1[a]	**2430**	-	-	-	This work
		1[a]	**2700**	-	-	-	2010
Fucα(1-2)Gal disaccharide		1[a]	**2500**	-	-	-	This work
H-type 1 trisaccharide		0.86 ± 0.10	**25.4 ± 4.5**	26.3 ± 0.4	41.1 ± 2.7	14.8 ± 2.3	This work
H-type 1 tetrasaccharide		1.01	**56.6**	24.3	37.5	13.2	2020
		0.93 ± 0.02	**77.2 ± 1.5**	23.5 ± 0.2	23.0 ± 0.3	−0.5	2010
Lewis y pentasaccharide		0.99	**52.6**	24.4	43.3	18.8	2020
		0.98 ± 0.03	**53.9 ± 2.9**	24.4 ± 0.2	34.9 ± 0.3	10.5	2010
Globo H (H-type 3) hexasaccharide		0.83 ± 0.06	**26.1 ± 1.7**	26.1 ± 0.2	46.1 ± 3.9	20.1	2020

Table 4.3. Affinity and thermodynamic measurements for carbohydrate ligands of BC2L-C-N_{ter}. Experiments performed at 25 °C. Averages and experimental errors for at least two independent measurements. [a]Stoechiometry fixed during the fitting procedure: thermodynamic values cannot be assessed. Adapted from Šulák and co-workers (2010) and Bermeo and co-workers (2020).[134, 165]

4.4. Outlook

The first part of this project had two main objectives: to produce a new, reliable protein construct of BC2L-C-N_{ter} and to gain as much insight as possible in regard to the protein and its binding interactions. The success of the first objective led the way towards the second: the new recombinant construct proved to be remarkable stable without drawbacks in terms of lectin function. Therefore, it is suitable to establish reliable protocols for biophysical evaluation by techniques such as ITC, SPR, etc. Furthermore, the construct rBC2L-CN2 opened the door for crystallographic study of its interactions with human epitopes such as the histo-blood groups. It is worth to mention that this construct is able to generate sizeable crystals under extremely simple conditions, in a timely and consistent manner. These crystals, either in apo-form or in complex, diffract at high resolution (1.9 Å and higher).[167]

The three new crystal structures examined thus far confirm established knowledge and provide new information. On one hand, the novel fucose binding mode described in the seminal study is consistently observed across structures, down to the water molecules present in the CRD. This accounts for the fucoside selectivity and the millimolar affinity measured for monovalent fucosides. On the other hand, the micromolar affinity for oligosaccharides, known since 2010, was finally rationalized to the third sugar unit. This

rationalization was the result of affinity investigation through ITC and structural study of the binding interactions.

Noteworthy, the spatial position of the larger ligands was found to be considerably solvent-exposed and the key residues and water molecules involved in the binding were identified. This is a crucial step for the rational design of synthetic ligands. Two different strategies are possible: to emulate the binding mode by targeting the identified key residues or to try to tap into vicinal sites untouched by the oligosaccharide ligands. The two strategies can eventually be combined but first need to be validated on their own.

5. DESIGN AND SYNTHESIS OF ANTAGONISTS

5.1. Summary

The work presented in this part was performed starting in 2019, over the period of 14 months at the University of Milan. It had two objectives: to attempt the rational design of antagonists towards the now well-characterized target and to establish and validate synthetic routes towards the first final molecules. Bachelor and Master students N. Quadrio, D. Ruggeri and D. Lanaro were mentored during or after the period at University of Milan, and contributed to the synthetic project. Given the time constraint, the main interest was to establish reliable methodology that could be replicated and continued in Milan after the transfer to Grenoble.

With all the necessary tools at hand, the rational design of BC2L-C-N_{ter} could start. On one hand, the crystallographic study showed that the most conserved interactions were the ones established by the terminal fucoside unit. On the other hand, computational methods were available through the collaboration with Kanhaya Lal, PhD student from PhD4GlycoDrug.

Thanks to *in silico* screening of fragments, a small library of fragment structures was obtained, that could fill a site vicinal to the CRD of the lectin. The choice was made to connect these fragments to a monovalent fucose ring, generating glycomimetic structures with the potential to encompass both the CRD and its vicinal site. As a first step, linking functionalities were selected based on synthetic feasibility, directionality, and predicted distance from the sugar anomeric position to the docked fragments. Then, a synthetic scheme was drafted that would combine L-fucose and various ligand structures into bifunctional glycomimetics featuring different linking functions. To ease and maximize the output of final molecules, the synthetic route was devised to be modular and convergent. The synthetic project proceeded successfully, generating a panel of final molecules, ready to be evaluated against BC2L-C-N_{ter}. Alternatively, some branching strategies were explored, which could be beneficial in future ventures related to multivalency.

The following sub-sections present the rational design of ligands, the drafting of a modular synthetic route and its implementation, followed by branching strategies. Taken together, the design and synthesis of potential BC2L-C-N_{ter} ligands was successful and established a solid

framework to continue producing new synthetic molecules with relative ease. Experimental procedures and characterizations can be found in the **Experimental Section** (**APPENDIX 8.3**).

5.2. Design of monovalent fucoside antagonists

A recently published article describes the efforts of K. Lal towards his project: "Design of lectin antagonists through fragment-based screening and molecular modelling".[167] The article describes the first computational steps towards BC2L-C-N_{ter} ligand design, and is available in the **APPENDIX 8.2: Prediction and Validation of a Druggable Site on Virulence Factor of Drug Resistant *Burkholderia cenocepacia***. Indeed, virtual screening has recently been successfully used to develop inhibitors for bacterial lectins.[169] Thus, the 2010 crystal structure of the BC2L-C-nt/αMe-Seleno-L-Fucoside complex was used to screen the vicinity of the CRD for potential 'ligandable' vicinal sites.[170] This preliminary analysis was performed with the informatic tool SiteMap and provided three main outputs shown in **Figure 5.1** as sites X, Y, and Z.[171]

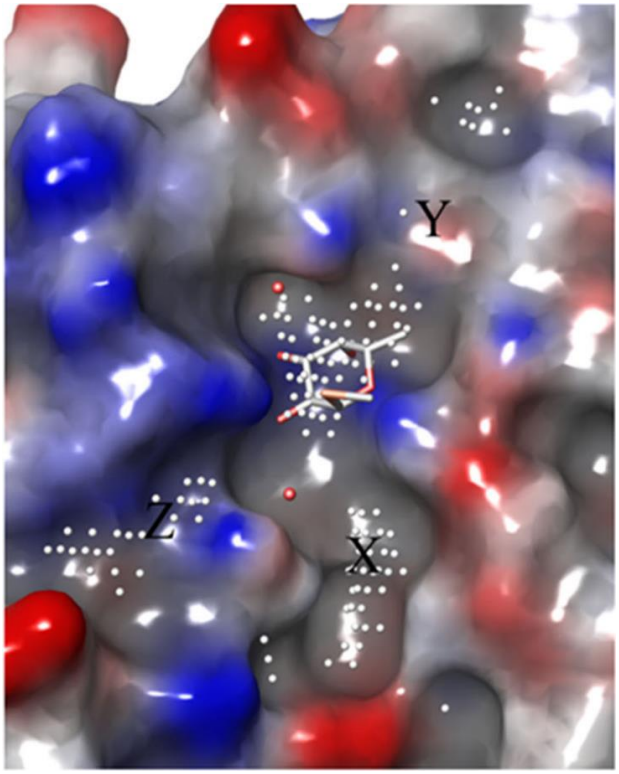


Figure 5.1. Output from SiteMap analysis of BC2L-C-N_{ter}'s CRD: clusters of points reveal vicinal sites suitable for fragment binding. Adapted from Lal and co-workers (2021).[167]

Among the sites considered, two were solvent-exposed (Y and Z). Site Y corresponds to the binding site of oligosaccharide ligands and Z was its opposite across the CRD. Although

shallow binding sites are a staple of lectins, they often accommodate ligands with low affinity and are not the best suited for development of ligands that aim for high affinity and high selectivity.[46, 172] Site X appeared to be more promising: depth and balance of polar and hydrophobic residues. This crevice, which follows the protomeric binding interface, was ideal for fragment screening. A third parameter made site X particularly interesting: its orientation in relation to the fucose binding site, which suggests the fragments could be connected to the anomeric position (C1) of fucose.

Hence, a library of 2000 fragments was screened against site X and docked in the presence of the monosaccharide using Glide.[173] The structures susceptible to bind the vicinal site were ranked and shortlisted, leading to the purchase of 12 fragments (see **Figure 5.2**), which were later validated by an array of biophysical techniques as described by Lal and co-workers (2021).[167]

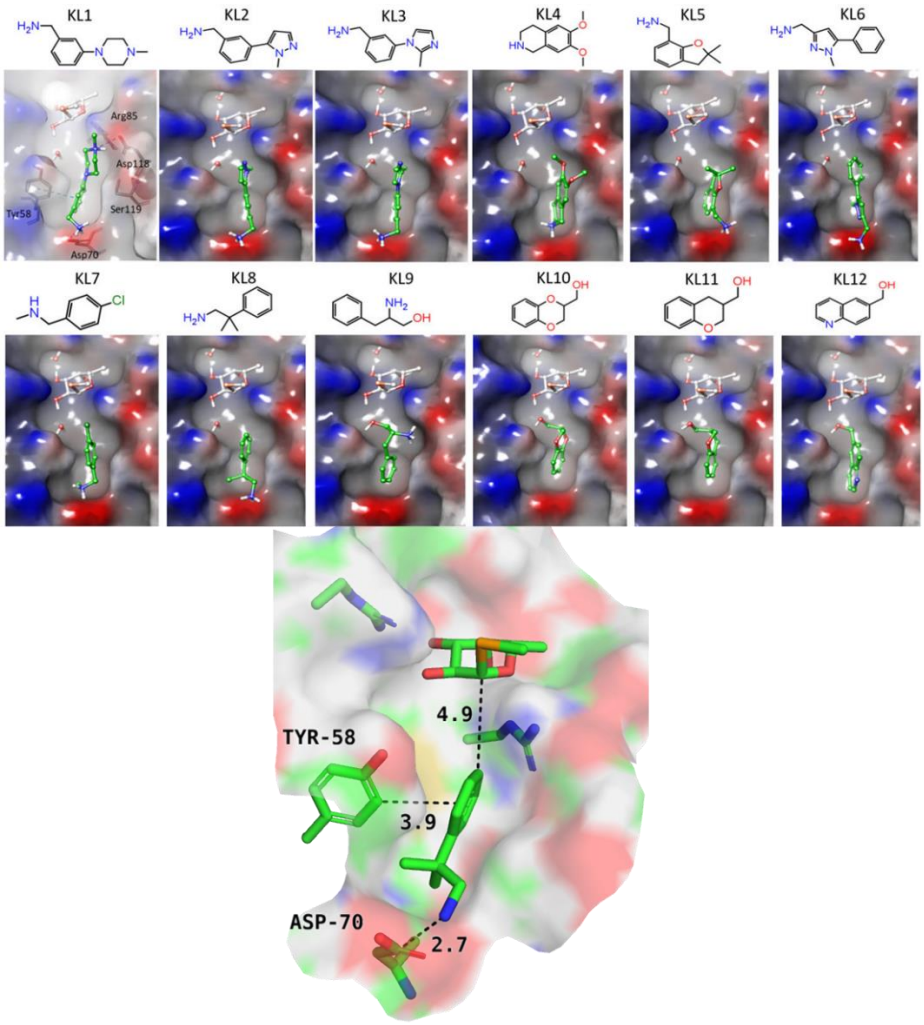


Figure 5.2. Right: Binding poses for the top ranked fragments (**KL01 - KL12**) predicted by docking studies at site X. **Left:** Focus on **KL08**: distance to fucoside and predicted interactions. Distances in Å. Adapted from Lal and co-workers (2021).[167]

These structures typically featured amino-substituted aromatic (hetero)cycles. The combination of aromatic moieties and polar/charged substituents targeted specific residues such as **Tyr58** and **Asp70**. Moreover, the fragments docked in proximity to the CRD: a few angstroms away (3 - 6 Å) from the anomeric carbon of the fucoside. This proved to be an advantage later on: carbohydrate functionalization through the anomeric position is easy and versatile.

Indeed, the general idea for ligand design was to link the available fragments to a fucoside core to generate bifunctional molecules able to bind to the CRD and site X simultaneously. The concept of targeting neighboring areas of a CRD has been a staple of glycomimetic design, often targeting hydrophobic residues to produce glycomimetic structures that counterbalance the inherent hydrophilicity of sugars. Naturally, this concept also allows enhancement of selectivity and affinity.[56] Thus the designed fucosides were tailored with the goal of replicating the observed and simulated binding poses without significant steric clash (see **Figure 5.3**).

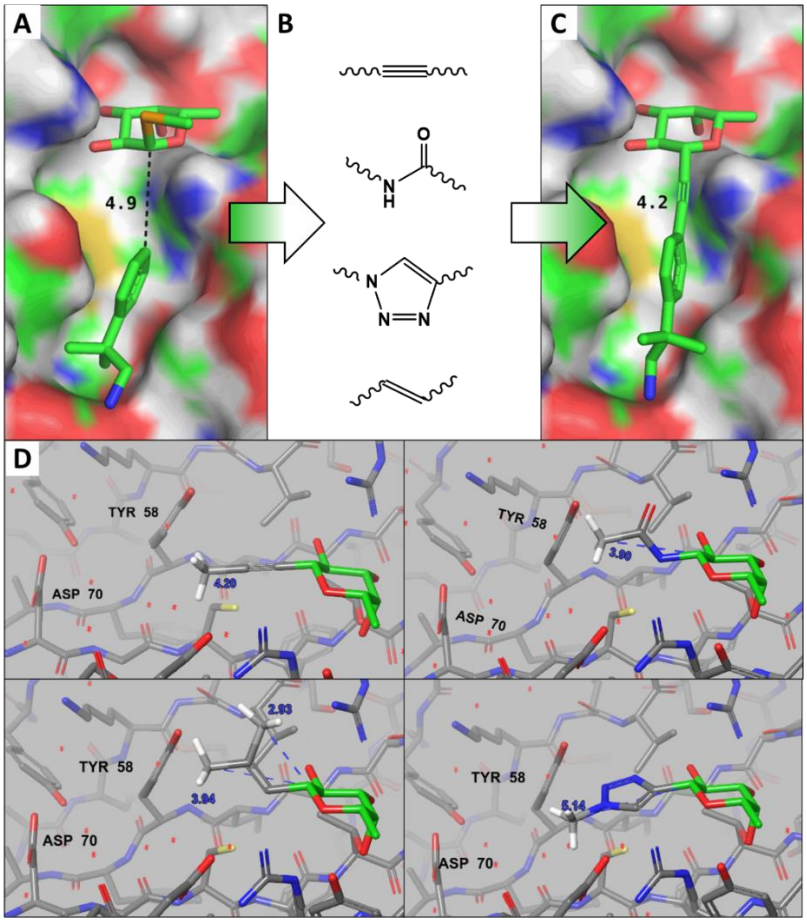


Figure 5.3. Ligand design strategy and linkages considered. **A:** Example of a fragment screening hit: **KL08**. **B:** Chemical linkages considered. **C:** Example of a docked final molecule. **D:** Visualization of the linkages reaching into vicinal site X. Distances from anomeric carbon in Å (4.20, 3.90, 3.94, 5.14).

For this, using a variety of chemical linkages would be advantageous: factors such as rigidity, orientation, length and general size could be modulated by choosing different linkers. Finally, to remain close to the "glycodrug" goal, unnatural linkages were preferred to the classic O-glycosidic bonds which, although easily applicable in this particular case, represent a metabolic and hydrolytic soft-spot.[56]

Consequently, a set of functions were chosen while keeping synthetic feasibility in mind. Amide, triazole, alkyne and alkene functions were considered to be chemically simple, yet robust and broadly applicable, and presenting different characteristics in terms of bridging length, angle, flexibility, bulkiness, polarity and metabolic stability. Among these, the alkyne function was particularly interesting: the β-fucosylacetylene had the exact orientation needed and an acceptable length (4.2 Å, see **Figure 5.3**), compatible with the predicted docking poses in site X. Alternatively, the amide bond offered polar surfaces to interact with the nearby, crystallographically conserved water molecule **W2** (see **Figure 5.4**). Other linkages that seemed accessible at this stage were the alkene bond, through its E-position and the long-reaching triazole function.

It is worth to mention that all linkages were considered as β-oriented from the anomeric carbon: the orientation perfectly matched the vicinal site. The first generation of antagonists was thus designed as a panel of β-C- and β-N-fucosides targeting the CRD of BC2L-C-N_{ter} and its vicinal site X. To complement the design of these ligands, docking experiments were performed by K. Lal for some of the expected final molecules, providing information on how they are expected to fit within the target (see **Figure 5.4**).

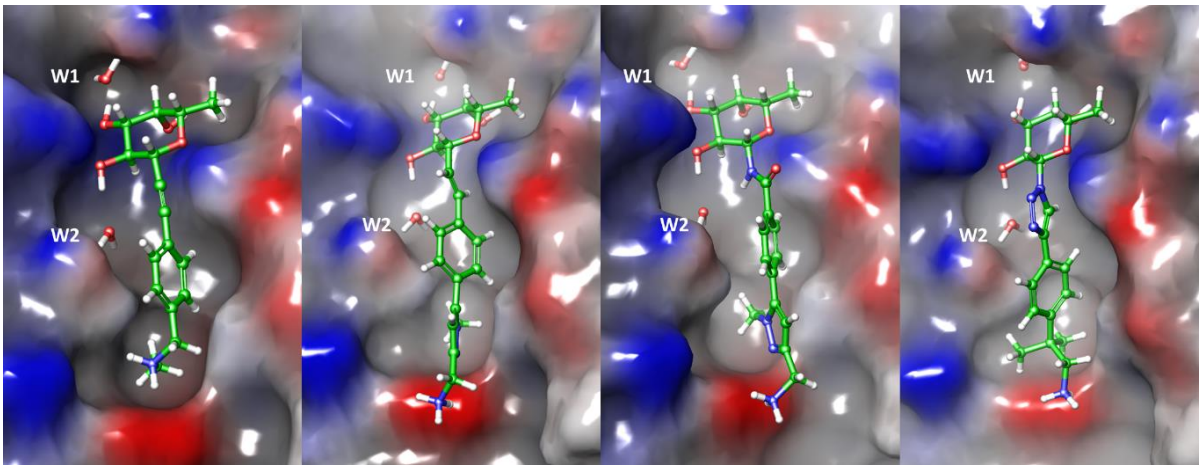


Figure 5.4. Examples final molecules docked on BC2L-C-N_{ter}. **From left to right:** alkyne-bound **KL07**, alkene-bound **KL06**, amide-bound **KL06**, triazole-bound **KL08**.

As a final note, it's worth of mention that the computational work of K. Lal was validated later on, when a crystal structure was obtained featuring fragment **KL03** in its predicted binding pose (see **APPENDIX 8.2**).[167] The preliminary ligand design was, thus, finished and the next step involved the drafting of synthetic routes towards the designed molecules in a time-effective and straightforward manner.

5.3. Modular synthesis of C- and N-fucoside glycomimetics

The synthetic work of this project consisted in linking various fragments to a fucoside core in order to produce a panel of glycomimetics. Two interrogations surfaced early on:

- Which final molecule should be pursued first?
- If the synthetic route is successful, can it be adapted to subsequent structures?

Possible answers to the first question were to prioritize the fragment with the best score from the screening's scoring function or to choose the fragment which was easiest to functionalize. However, both options were flawed. On one hand, a scoring function should never be used as a predictor of affinity and a high score doesn't denote a real advantage if all scores are in the same range. On the other hand, basing a synthetic route on a handpicked achievable structure certainly works but the resulting route is difficult to adapt to subsequent structures that are bound to be more challenging. This would lead to establishing new synthetic routes for every final molecule, which didn't align with the time constraints.

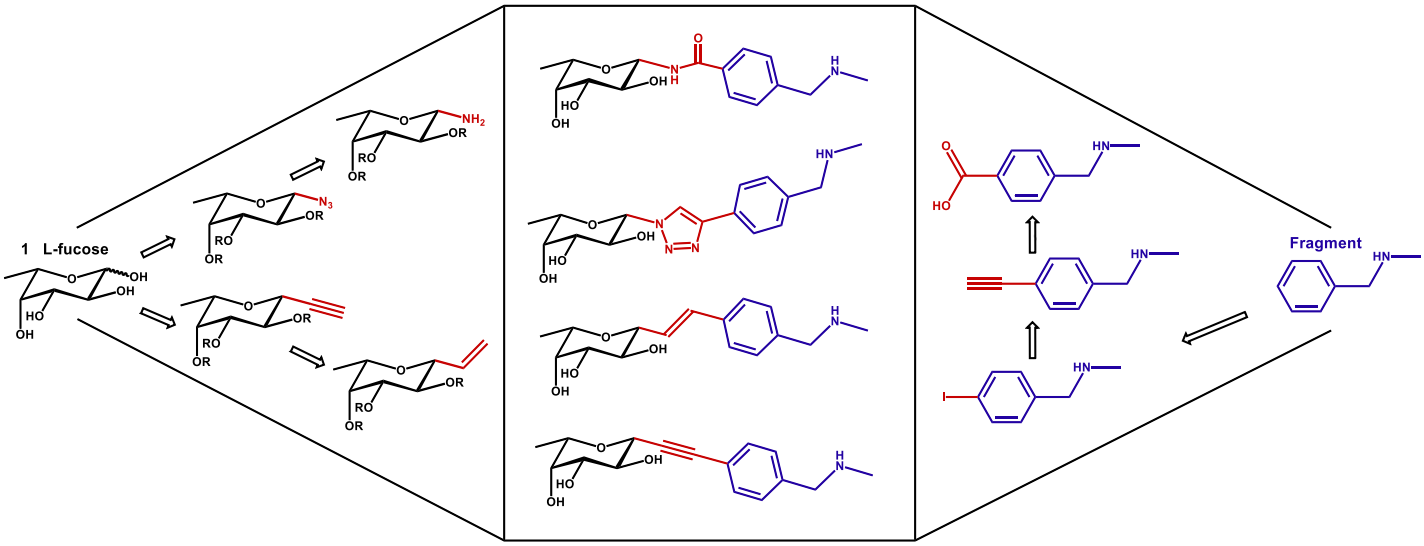
The solution was to establish a modular synthesis: in this way synthetic intermediates needed for one final molecule can also be used for a different one, minimizing the number of unique steps and allowing for quick large scale synthesis of a handful on 'milestone' intermediates. Since the constant motif was the fucose core, the logical choice was made to implement the different linking functions on this moiety and leave the attachment to fragments for the later steps. However, this meant that the late-stage coupling procedures needed to be broadly applicable in order to accommodate the structural diversity of the purchased fragments and any other fragments subsequently added to the screen.

Thus, the synthetic route towards the designed fucomimetics was drafted to satisfy two requirements: (1) modularity, allowing for all designed final molecules to be synthesized from the same building blocks; (2) robust and reliable coupling procedures. Naturally, the

advantage of such synthetic route is that, as long as the required 'milestone' intermediates are obtained, the set of final molecules should be quickly accessed (see **Scheme 5.1**). A rapidly growing set of derivatives would open the way to empirically compare the contributions of different fragments and linkages and establish structure-activity relationships (SAR). Therefore, a modular framework for synthesis of BC2L-C-N_{ter} would allow for prioritization, essential in the early stages of a project such as this one.

As it transpires from **Scheme 5.1**, the modular strategy required fragments to feature a CRD-adjacent position that could be easily iodinated. This was not the case: the intricacies of ligand functionalization are discussed in the following sub-section (**5.4 Fragment functionalization strategy**). For the general case, iodination was considered to be broadly applicable as a functionalization strategy. Consequently, the 'milestone' intermediates necessary for the modular synthesis are three: an iodinated fragment, a β-azidofucoside and a β-fucosylacetylene. On the other hand, the coupling procedures necessary to attain the final molecules are staples of organic synthesis: the amide bond coupling, the 'click' copper-catalysed azide–alkyne cycloaddition (CuAAC) and the Sonogashira and Heck coupling reactions.[174-178]

Scheme 5.1. Modular synthesis towards β-C- and β-N-fucosides exemplified for fragment **KL07**.



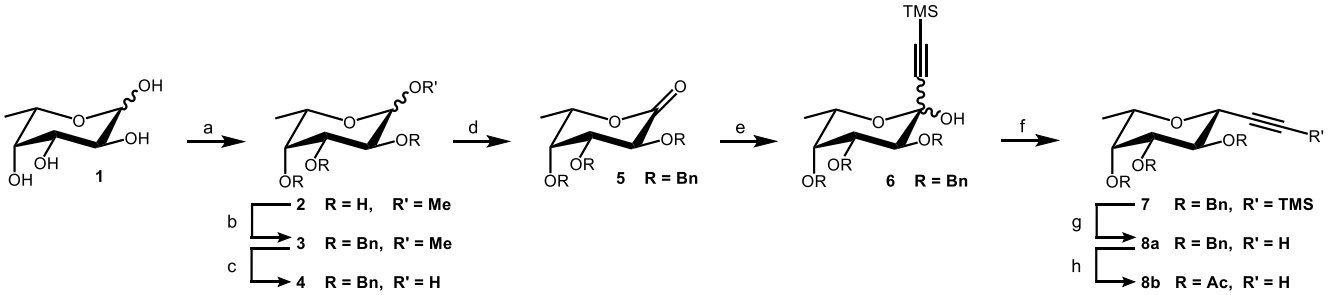
a. Synthesis of β-C-fucosides

The first intermediate pursued was the β-C-fucosylacetylene: it opened the way not only to alkyne-bound molecules but could also lead to the alkene and triazole linkages. It also

represented synthetic novelty, since β-C-fucosylacetylenes had never been prepared as building blocks. On the other hand gluco-, manno- and galacto-equivalents have been produced β-selectively, laying the groundwork for this attempt.[179-181]

The synthetic route towards intermediates **8a** and **8b** is described in **Scheme 5.2**: starting from L-fucose **1**, a series of protection and deprotection reactions led to molecule **4**, appropriately protected in all positions but the anomeric one. Initially, these steps were hindered by the necessity to purify the relevant α- and β-pyranose forms from the undesired furanose forms. Indeed, the starting material **1** and the methyl fucoside **2** exist as an equilibrium of these forms. A first solution to ensure pyranose purity was to recrystallize the crude of the anomeric methylation reaction: the α-form of product **2** (methyl α-L-fucopyranoside) was obtained as pure crystals with maximum yields of 56% (conditions detailed in **Scheme 5.2**). Indeed, as described by Mowery in 1975, an equilibrium between four forms also exists during the Fischer methylation conditions.[182] Along with the recrystallization, a recycling protocol reinstated Fischer reaction conditions (**a** in **Scheme 5.2**) on the mother liquor in order to re-equilibrate the amount of methyl α-L-fucopyranoside to over 50%, as was observed by the anomeric signals by ^{1}H NMR (D_2O: in order for α/β-pyranose and α/β-furanose: 4.77, 4.31, 4.89, 4.93 ppm). Thus, consecutive recrystallization and recycling allowed to maximise purity and yields at the onset of the synthetic route.

Scheme 5.2. Synthetic route towards the milestone β-C-fucosylacetylenes **8a** and **8b***



***Reagents and conditions: a.** Amberlite® IR_{120} H^+, MeOH, 65 °C, 56%; **b.** BnBr, KOH, Tol, 111 °C, 80%; **c.** HCl, AcOH, 118 °C, 78%; **d.** DMP, DCM, rt, 79%; **or** I_2, K_2CO_3, DCM, rt, 75%; **e.** TMS-acetylene, nBuLi, $CeCl_3$, THF, -78 °C, 87%; **f.** Et_3SiH, $BF_3{\cdot}Et_2O$, CH_3CN/DCM, -10 °C, 86%; **g.** NaOH, MeOH/DCM, rt, 99%; **h.** TMSOTf, Ac_2O, rt, 61%. Re-crystallization of **2**: dissolution in EtOAc to *ca.* 55 mg/mL, reflux (77 °C) and cooling to 5 °C, then filtering.

The second solution was more straightforward: commercially available methyl α-L-fucopyranoside was bought and used for the rest of the synthetic project. Back to 2,3,4-tri-O-benzyl-L-fucopyranose **4**, two protocols were used to oxidize the anomeric position and obtain the fuconolactone **5**. A protocol developed by Koch and co-workers (2003) for a D-

galactoside was followed, which used Dess-Martin periodinane (DMP) and resulted in good conversion but also in difficult separation of **5** from the reaction by-products.[183] The second set of conditions, developed by Wei and co-workers (2015), involved iodine and led to similar conversion with easier purification, becoming the method of choice with good yields (75%).[184]

From **5**, an organocerium reaction allowed to install the acetylene moiety resulting in molecule **6**.[179, 181] This moisture-sensitive reaction was particularly challenging in terms of experimental set-up: on one hand, $CeCl_3$ was dried at 140 °C under vacuum, then suspended in freshly-distilled THF at 0 °C and cooled to -78 °C under argon; on the other hand, TMS-acetylene in freshly-distilled THF was reacted with excess of n-BuLi, prior to transfer via cannula to the cerium suspension, all at -78°C under argon. Finally, cannula addition of cold lactone **6** in freshly-distilled THF allowed the functionalization to proceed for 2 hours at -78 °C, under argon. This reaction was ultimately mastered and led to good yields even in gram-scale. A fortuitous event was observed for the large-scale version of the reaction: the additional equivalents of nBuLi necessary to bring the reaction to completion led to the deprotection of the alkyne, providing a mixture of the intended **6** and its TMS-free version. This helpful effect can be optimized in the future to shorten the route by one step.

Up to this point, α/β mixtures are unavoidable but also harmless. This is because the following step ensures the β-selectivity of the route: β-stereoselective deoxygenation converts anomeric mixture **6** into β-fucosylacetylene **7**. This effect is achieved by the organosilane-boron trifluoride reducing system: the hydroxyl group is activated by a Lewis acid (boron trifluoride etherate) and becomes a leaving group, generating the oxocarbenium ion.[185] Subsequently, a nucleophilic attack by the mild hydride source (silane) axially delivers a hydrogen to the intermediate.

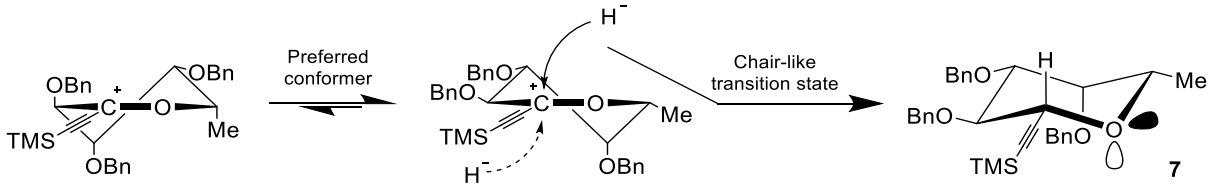


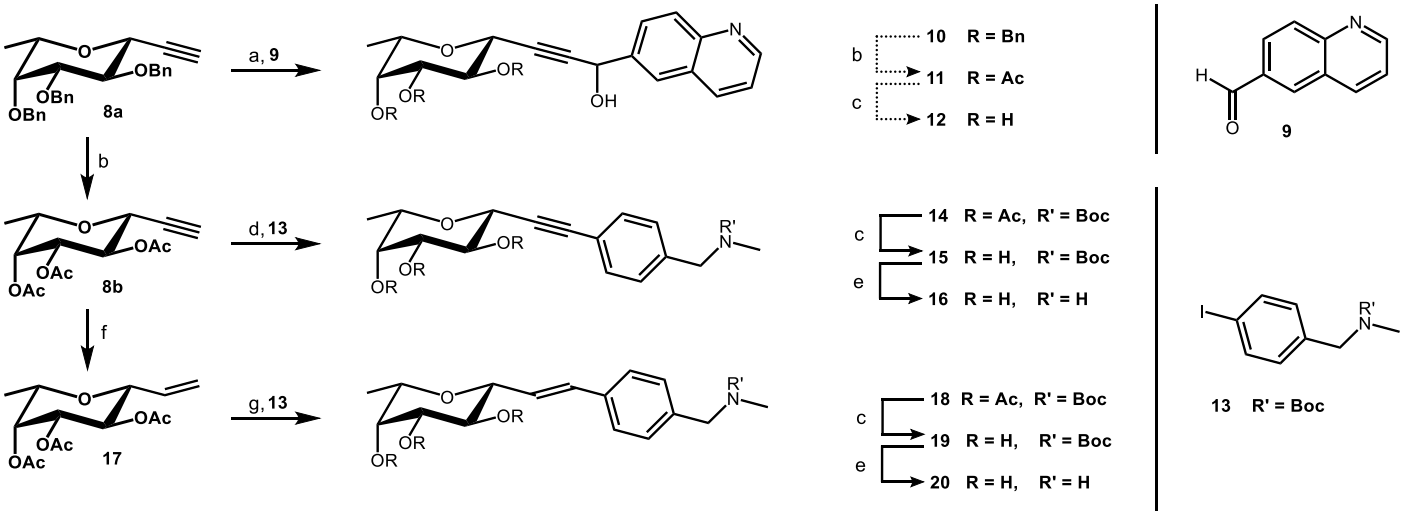
Figure 5.5. Diastereomeric half-chair conformers of oxocarbenium ions undergo nucleophilic attack by a putative hydride to follow the favored chair-like transition state and yield β-fucosylacetylene **7**.

As detailed in **Figure 5.5**, this α-selective attack is likely directed by the antiperiplanar electron pair of the ring oxygen, in what could be called an "anomeric effect".[186-188] Nevertheless, it is also explained by the favoured transition from the preferred half-chair conformer to a chair-

like transition state, as opposed to a high-energy twist-boat. A similar case has been described for glucal oxocarbenium ions.[189]

Onwards, having deprotected the alkyne on **7** to obtain the milestone intermediate **8a**, the synthetic route forked into three branches. The first branch (middle fork of **Scheme 5.3**) dealt with the removal of the benzyl protecting groups: after the organocerium reaction, they were no longer necessary and could be removed. However, classical deprotection conditions involved the use of palladium on carbon as catalyst and hydrogen gas, which were incompatible with the substrate. To avoid the accidental reduction of the triple bond, the benzyl groups were removed using an acetolysis reaction (TMSOTf, Ac_2O), providing **8b** with 61% yield. The protocol used required long reaction times: as described for similar substrates, the last benzyl group, probably in position 3, needed between 24 and 48 h to be replaced.[181] The O-acetyl fucoside **8b** was employed as seen in the central fork of **Scheme 5.3**: Sonogashira coupling reaction ($Pd(PPh_3)_4$, CuI, iodinated fragment, Piperidine), followed by Zemplén deprotection and Boc-removal. These reactions ran smoothly with yields over 80%, providing final molecules as TFA salts: **16** and **27**, featuring fragments **KL07** and **KL08,** respectively (details in **Table 5.1**).

Scheme 5.3 Synthesis of β-C-fucoside final molecules, exemplified for fragments **KL12** and **KL07***



***Reagents and conditions: a.** LDA, **aldehyde 9**, THF, -20 °C, 72%; **b.** TMSOTf, Ac_2O, rt. **c.** MeONa, MeOH, rt, quant.; **d.** Sonogashira: $Pd(PPh_3)_4$, CuI, **fragment 17**, Piperidine, 80 °C, 81%; **e.** TFA, DCM, 0°C; **f.** Lindlar's Pd Catalyst, H_2, MeOH, 89%; **g.** Heck: $Pd(OAc)_2$, KCl, TBAB, K_2CO_3, $AgNO_3$, DMF, 100 °C, 81%. Omitted yields reported in **Table 5.1**.

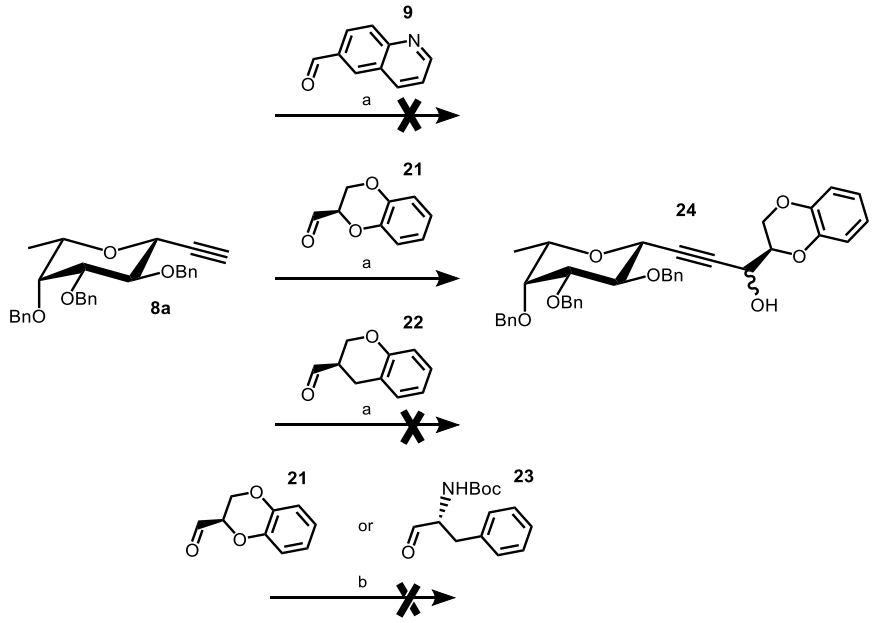
The next branch (bottom fork of **Scheme 5.3**) went similarly. After a high-yielding reduction of alkyne **8b** (Lindlar's Catalyst, H_2, MeOH), the free alkene **17** was used for Heck coupling.[190] A coupling protocol adapted for vinyl sugars ($Pd(OAc)_2$, KCl, TBAB, K_2CO_3, AgNO3, DMF) was employed with the same coupling partner **13** and provided coupled molecule **18** in good yields

(81%).[191] The necessary deprotections (Zemplén and Boc-removal) followed to afford final molecule **20** as the TFA salt, thus validating this branch and the alkene linkage.

The last branching point was the hardest to establish. It led to bifunctional molecules featuring a propargylic alcohol moiety (top fork of **Scheme 5.3**). Indeed, some of the 12 fragment hits from the computational screen bore a hydroxyl group directed towards the CRD, predicted to replace the crystallographically conserved water molecule **W2** (see **Figure 5.2**, fragments **KL09 - 12**).[167] Entropically speaking, successful replacement of an ordered water molecule while maintaining its interactions can translate into a considerable affinity gain.[166] Consequently, we aimed to validate this route with at least one fragment.

Test reactions detailed in **Scheme 5.4** were performed on small scale with aldehyde versions of fragments **KL10 - 12**, following a protocol adapted from Dondoni and co-workers (2002).[192] The alkyne anion was generated with LiHMDS at -20 °C and the THF-dissolved aldehyde was added at -45 °C. Since two out of three reactions did not proceed (unreacted starting materials), it was clear that the conditions needed optimization in terms of base equivalents and setup. Nevertheless, the reaction involving aldehyde **21** proceeded, albeit its crude was difficult to separate, leading to very low yields after chromatography (10%).

Scheme 5.4. First attempts to produce ligands featuring the propargylic alcohol moiety*



***Reagents and conditions: a.** LiHMDS, **aldehyde**, THF, -45 °C, 10%; **b.** $Zn(OTf)_2$, (+)-N-methyl ephedrine, Et_3N, **aldehyde**, Tol, rt or 111 °C.

The few milligrams obtained of product **24** seemed to indicate a successful coupling. Although impure, the signals for both moieties were observed by ^{1}H NMR ($CDCl_3$): among others, 6.79

- 6.92 and 7.26 - 7.40 ppm for the fragment and benzyl groups, respectively, with no trace of alkyne or aldehyde protons (2.51 and 9.77 ppm, respectively). The reaction was deemed feasible and improvable but wasn't optimised as such: it surfaced that a better protocol could be attempted. The work of Carreira and co-workers (2000 and 2002) described milder conditions that could even resolve the expected stereoisomeric mixture.[193-194] Nevertheless, small-scale attempts to use this methodology on the fucoside substrates did not provide any results (conditions **b.** in **Scheme 5.4**). Immediate optimization wasn't performed right away: time constraints shifted the focus to other activities.

It was later on, at the hands of D. Lanaro, that this coupling reaction would be re-visited and validated (**Scheme 5.3**, upper fork): the 'Dondoni' conditions were optimized, shifting to freshly prepared Lithium diisopropylamide (LDA). The reaction was attempted on aldehyde **9**, which had proven to be the better choice. Indeed, searching the literature revealed the poor stability of aldehyde **21**, which partially explained the difficulty to unambiguously characterize **24**.[195] On the contrary, product **10** could be clearly identified by MS and ^{1}H NMR ($CDCl_3$) as a diastereoisomeric mixture: some signals, such as which of *H-2* were clearly doubled (δ = 5.70 and 5.69 ppm). The conditions for acetolysis and full deprotection to afford **12** have been established and will allow this avenue to be fully explored in the follow-up of the project. Thus, couplings towards three types of β-C-fucoside ligands were drafted and validated as summarized in **Table 5.1**.

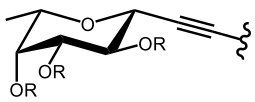
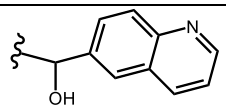
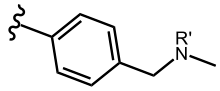
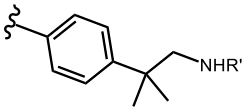
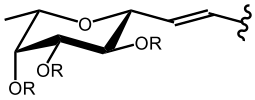
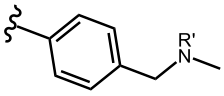
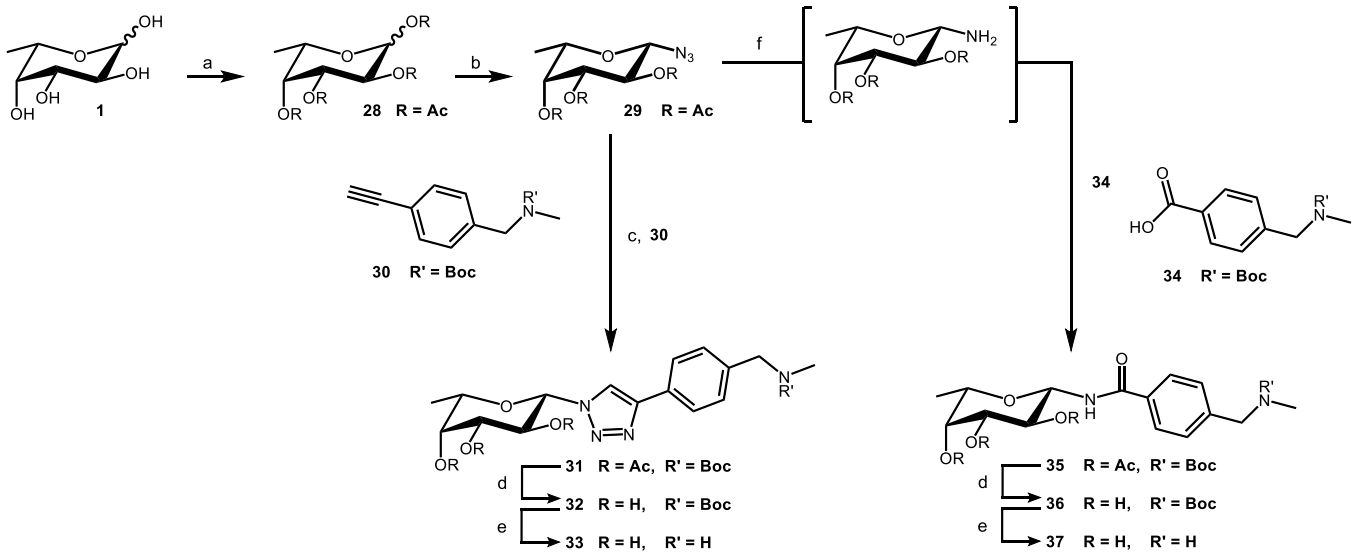
Fucoside moiety	Fragment moiety	R = Ac R' = Boc	R = H R' = Boc	R = H R' = H
		**10**[a] 72%	**11**[b] -	**12** -
		**14** 85%	**15** 97%	**16** quant.
		**25** 85%	**26** quant.	**27** quant.
		**18** 81%	**19** 49%	**20** quant.

Table 5.1. Panel of β-C-fucoside final molecules. Reagents and conditions as shown in **Scheme 5.2** and **5.4.** [a]R = Bn. [b]R = Ac.

b. Synthesis of β-N-fucosides

At the same time as the route towards the β-fucosylacetylenes provided the first final molecules, D. Ruggeri was entrusted with the task of implementing the synthetic route that had been drafted towards the other milestone intermediate: the β-azidofucoside. **Scheme 5.5** outlines the synthetic route towards intermediate , which leads to amide- and triazole-bound final molecules. Starting from L-fucose **1**, per-acetylation was followed by treatment by TMS-N_3 promoted by $SnCl_4$, providing target **29** in good yields (67% over two steps). The observed β-selectivity (α/β ratio 9:91) is explained by the participation of the neighbouring C2-bound acetyl, which forms an acyloxonium ion prior to the nucleophilic attack of the azide anion.[196-197]

Scheme 5.5. Synthesis of β-N-fucosides exemplified for fragment **KL07***

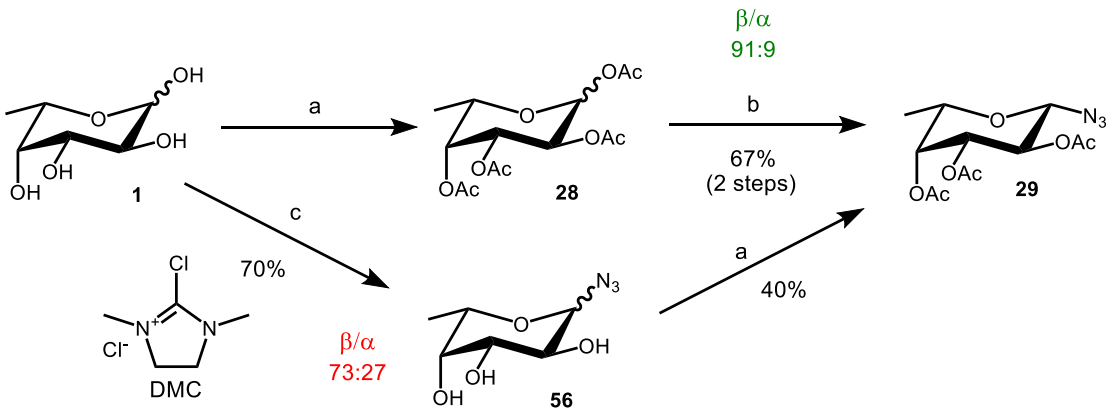


***Reagents and conditions: a.** Ac_2O, Pyr, rt; **b.** TMS-N_3, $SnCl_4$, DCM, 0 °C, 67% (over 2 steps); **c.** $CuSO_4 \cdot H_2O$, Na-Ascorbate, **alkyne 30**, MeOH, rt, quant. **d.** MeONa, MeOH, rt; **or:** NH_2Me, EtOH, rt; **e.** TFA, DCM, 0 °C; **f.** PMe_3, DCM, rt; **then: carboxylic acid 34**, HATU, DIPEA, DCM, rt; **or:** H_2, Pd/C, MeOH, rt; **then: carboxylic acid 34**, HATU, DIPEA, DCM, rt. Omitted yields reported in **Table 5.2**.

It is worth to mention that an alternative route towards the β-azidofucoside was explored: Tanaka and co-workers (2009) described a method to obtain β-azidoglycosides from unprotected saccharides, which was later adapted to L-fucose.[198-199] As seen in **Scheme 5.6**, this procedure was attempted with 2-chloro-1,3-dimethylimidazolinium chloride (DMC) in hopes of avoiding the need for acetylation. However, the reaction was not fully β-selective and provided an α/β ratio of 27:73. Furthermore, the anomeric mixture **56** could not be separated using reverse phase chromatography and required acetylation of the

azidofucosides to be effective. Combining this problem with the lower yields obtained, this procedure was disfavoured in comparison to the original approach.

Scheme 5.6. Routes evaluated towards intermediate 30*



***Reagents and conditions: a.** Ac_2O, Pyr, rt; **b.** TMS-N_3, $SnCl_4$, DCM, 0 °C; **c.** DMC, NaN_3, Et_3N, H_2O, 0 °C, 70%.

With intermediate **29** at hand, coupling with an alkyne or a carboxylic partner led to the triazole- and amide-linked molecules. Under standard CuAAC conditions, azide **29** and alkyne **30** provided triazole product **31** in quantitative yields. Similar conditions led to molecule **38**, both underwent de-acetylation and Boc-removal affording triazole-linked final molecules with ease (see **Table 5.2**).

On the other hand, the amide bond was obtained by an adapted Staudinger ligation protocol.[200] The azide was reduced to the fucosylamine using PMe_3 and the pre-activated (HATU) carboxylic fragment was added onto the fucoside. This tandem protocol afforded bifunctional molecules such as **35** with moderate yields. Although it needed optimization, this protocol allowed to perform the coupling directly from the azide **29**, befitting the modular requisite. An alternative protocol towards these structures was attempted by reducing the azide using H_2/palladium, then using the fucosylamine in a classic amide coupling. This second protocol was successful but low yielding. Although the yields for amide couplings remained improvable, standard deprotections yielded the final molecules in high yields, as detailed in **Table 5.2**.

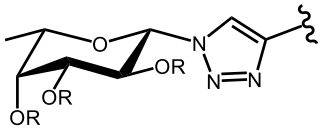
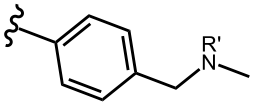
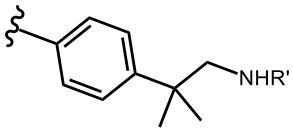
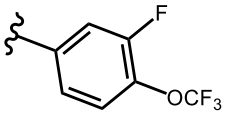
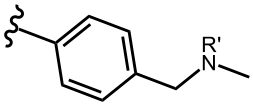
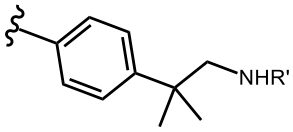
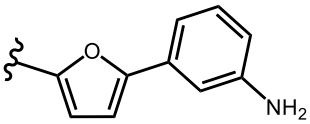
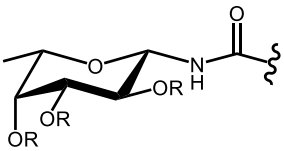
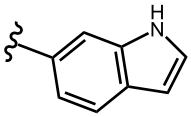
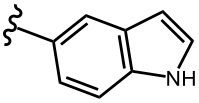
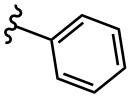
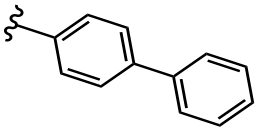
Fucoside moiety	Fragment moiety	R = Ac R' = Boc	R = H R' = Boc	R = H R' = H
		**31** quant.	**32** 60%[a]	**33** quant.
		**38** 86%	**39** 46%	**40** quant.
		**41** 56%	-	**42** quant.
		**35** 52%	**36** 62%	**37**[b] quant.
		**43** 45%	**44** 78%	**45** quant.
		**46** 23%	-	**47**[b] 82%
		**48** 29%	-	**49**[b] 92%
		**50** 13%	-	**51**[b] 94%
		**52** 44%	-	**53**[b] 98%
		**54** 34%	-	**55**[b] 90%

Table 5.2. Panel of β-N-fucoside final molecules. Reagents and conditions as shown in **Scheme 5.5**. [a]Over two steps. [b]Obtained from **46**, **48**, etc.

As transpires from **Table 5.2**, a handful of structures were used that didn't belong to the fragment screening. Some of these compounds were readily available and served to develop and validate the coupling protocols, yielding structures **41**, **52**, and **54**. The other three structures were deemed fit to target site X and were procured after validating them through docking of the hypothetic final molecules on the lectin target. Thus, the carboxylic acid leading to amide **46** was branded '**KL13**' and the carboxylic indoles **I1** and **I2** led to molecules **48** and **50**.

5.4. Fragment functionalization strategy

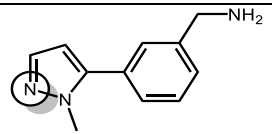
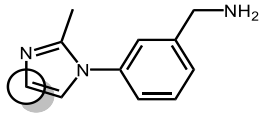
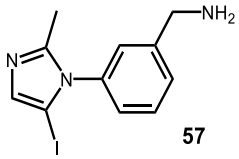
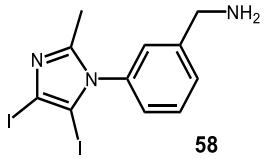
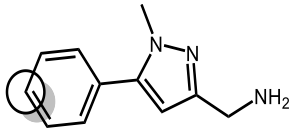
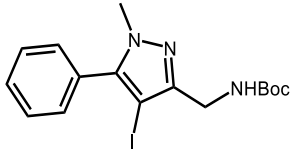
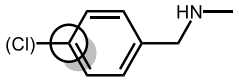
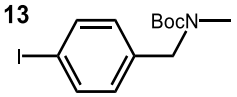
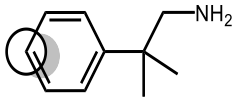
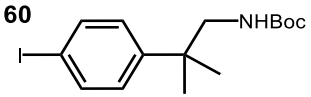
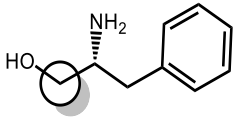
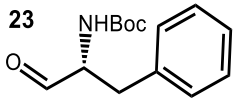
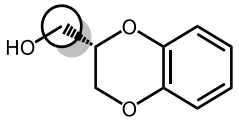
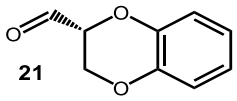
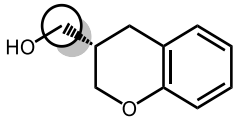
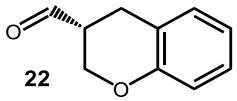
As mentioned in the previous section, one limiting factor to the modular synthetic framework was the functionalization of fragments. Indeed, each fragment required synthetic considerations adapted to its chemical features. Moreover, in order to accurately replicate the predicted binding pose, the possible attachment points were restricted to atoms neighbouring the fucose anomeric carbon with the right orientation. Indeed, a molecule that is 'pre-arranged' into its bioactive conformation benefits from entropic gains upon binding, leading to improved affinity.[46]

To optimize the strategies for fragment functionalization, the available fragments were separated in groups according to their features as seen in **Table 5.3**: (hetero)aromatics (**KL 2**, **3**, **6**, **7**, **8**), terminal alcohols (**KL 9 - 12**), and 'others' (**KL 1**, **4**, **5**). These screened fragments were complemented by a fourth group: carboxy-bearing fragments (**KL13**, **I1**, **I2**).

In order to fit the modular framework, iodinating the fragments at the desired attachment position was necessary. Alternatively, the primary alcohols could be reduced to aldehydes to afford secondary alcohols upon coupling, as described in the previous section. Since at least one of the purchased fragments contained a halogen at the appropriate position (**KL07**), the halogenation of (hetero)aromatic fragments was prioritized.

Attempts to couple the **KL07** in a Sonogashira protocol adapted for aryl chlorides didn't provide satisfactory results. A microwave procedure adapted from the work of Huang and co-workers (2008) led to 35% yield at best.[201] Therefore, it was interesting to generate the iodinated version of **KL07** starting from readily available N-methylbenzylamine. As **KL08** also required *para*-substitution, both were used to screen iodinating agents, resulting in the conditions described in **Table 5.3** (I_2, KIO_3, H_2SO_4, AcOH, 70 °C). Iodination of **KL08** led to molecule **60** with moderate *para*-selectivity. As seen by the relevant aromatic signals by ^{1}H NMR, the *para*/*meta* ratio was 87:13 ($CDCl_3$: 7.64 and 7.31 ppm, respectively), with trace impurities that may correspond to the *ortho* species. Attempt to increase the selectivity by reducing the reaction temperature failed due to the low reactivity of the substrate.

Further complications arose for the iodination of fragment **KL07** leading to **13**: a single spot on TLC was purified by chromatography and revealed to be a regioisomeric mixture by ^{1}H NMR. In this case, the *para*/*meta* ratio was 43:57 ($CDCl_3$: 7.63 and 7.56 ppm, respectively).

Group	Fragment	Structure (closest atom to fucose)	Conditions	Obtained molecule(s)	Yield
(Hetero) aromatics	KL02		Not yet attempted		
	KL03		a	 57 +  58	
	KL06		a, b	 59	64%‡
	KL07		a, b	13 	34%‡,#
	KL08		a, b	60 	64%‡,#
Terminal alcohols	KL09		b, c	23 	47%‡
	KL10		c	 21	54%
	KL11		c	 22	79%

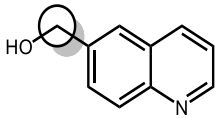
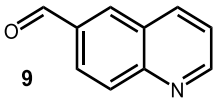
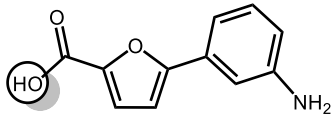
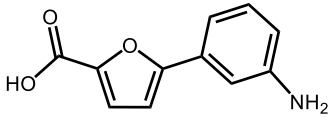
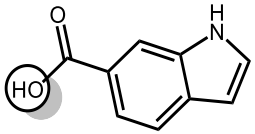
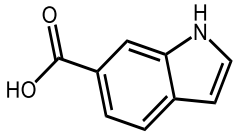
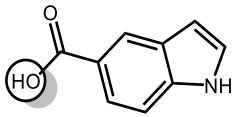
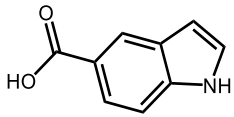
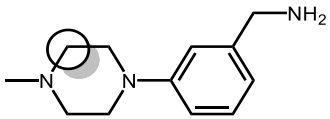
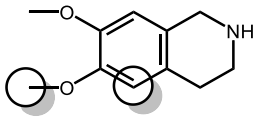
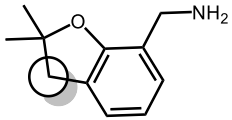
Group	Fragment	Structure (closest atom to fucose)	Conditions	Obtained molecule(s)	Yield
Terminal alcohols	KL12		**c**	 **9**	99%
Carboxylic acids	KL13		-		Commercially available
	I1		-		
	I2		-		
Others	KL01		Not yet considered for functionalization		
	KL04				
	KL05				

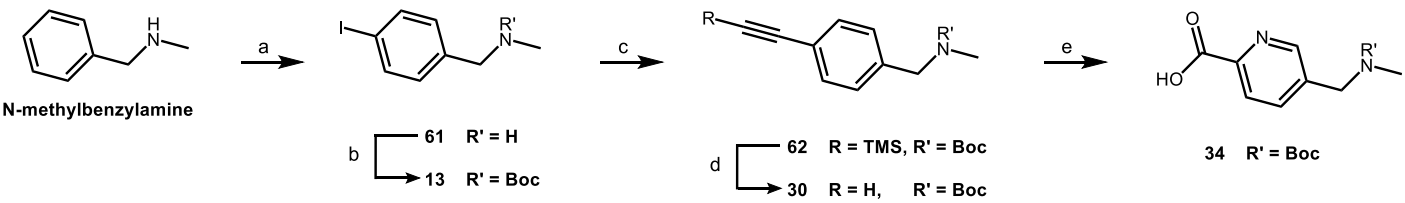
Table 5.3. Functionalization of fragments. **a.** I_2, KIO_3, H_2SO_4, AcOH, 70 °C. **b.** Boc_2O, DCM, rt **c.** DMP, DCM, 0 °C. ‡Yield over 2 steps. #Obtained as mixtures of regioisomers.

Therefore, these conditions weren't regioselective and regioisomeric mixtures **13** and **60**, which had respectively yielded 34% and 64% after two steps, were inseparable at this stage. The iodination of heterocyclic fragments **KL03** and **KL06** was even less successful: the activated heteroaromatic positions were iodinated preferentially, and hardly separable mixtures of regioisomers were obtained (see molecules **57** - **59**). Resulting from this impasse, it was decided that *de novo* synthesis of iodinated fragments would be necessary for some fragments (see below for **KL07**).

In the meantime, regioisomeric mixtures **13** and **60** were used as such to validate the rest of the synthetic route, in hopes of separating the *meta*-linked species at a further stage. Indeed, the final alkyne **16**, deriving from **13**, could be purified by HPLC, providing the intended structure for evaluation (HPLC: gradient of water to acetonitrile in a VP 250/21 column, see details in the **Experimental Section - 8.3**). On the contrary, HPLC purification of final molecule **27**, deriving from **60**, could not separate the *para*- and *meta*-linked species. As such, **27** remained contaminated by *ca.* 25% of *meta*-isomer. Since **27** ended up being a promising antagonist (a crystal structure was obtained featuring the *para*-isomer, see **Part 6**), a *de novo* synthesis of the fragment was drafted, which is ready to be implemented and leads unambiguously to *para*-linked **27** (see **Scheme 5.8**). In parallel to the iodination campaign, the aldehyde versions of alcohols **KL09 - 12** were obtained by oxidation without major problems. The improvable yields resulted from the difficult separation of Dess-Martin by-product but enough material was obtained before the conditions could be optimized.

Before *de novo* synthesis was attempted, it was relevant to use molecules **13** and **60** to validate as many synthetic routes as possible. Thus, the 'fragment' portion of the modular synthetic framework was devised to allow access to all final molecules from the iodinated 'milestone' intermediate. As shown in **Scheme 5.1**, fragments bearing alkyne and carboxylic acid functions were required. To achieve this, a series of functional group interconversions (FGIs) was employed, described in **Scheme 5.7**: a Sonogashira reaction installs the TMS-alkyne moiety (TMS-acetylene, $Pd(PPh_3)_4$, CuI, Et_3N, Toluene), which was subsequently deprotected (NaOH, MeOH/DCM) and oxidized to the carboxylic acid ($KMnO_4$, $NaHCO_3$, $tBuOH/H_2O$).[181, 202] Thanks to this strategy, any iodinated fragment can lead to 4 different final molecules featuring alkyne, alkene, amide, and triazole linkages. The strategy was successfully applied to obtain the **KL07** and **KL08** derivatives **30**, **34**, **65**, and **66,** with reasonable yields (see **Table 5.4**), leading to the synthesis of final molecules as detailed in the previous section.

Scheme 5.7. Fragment functionalization through FGIs*



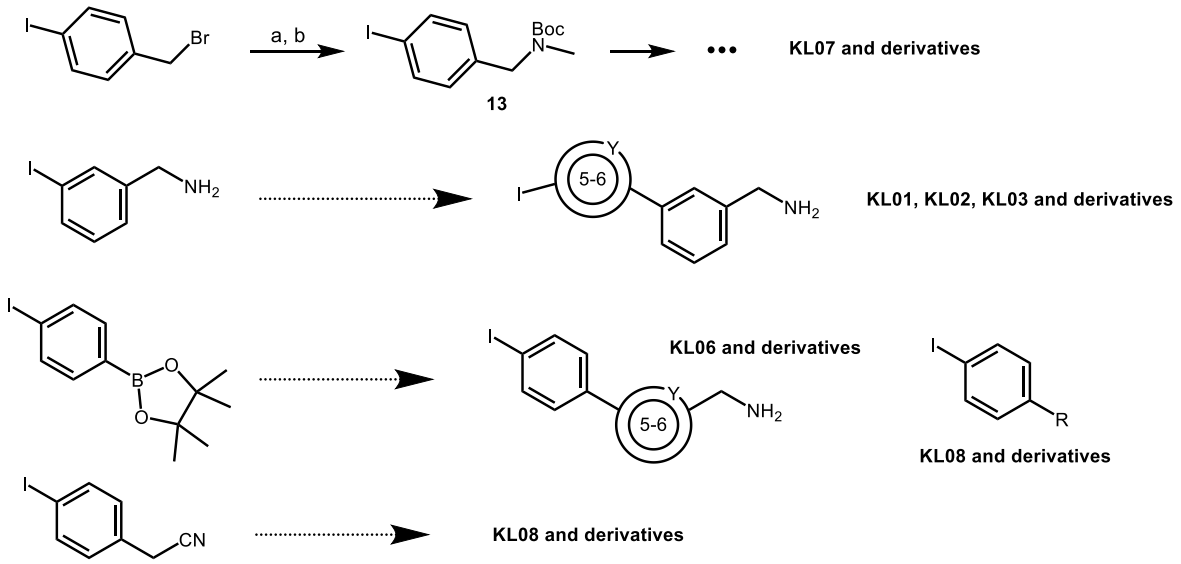
***Reagents and conditions: a.** I_2, KIO_3, H_2SO_4, AcOH, 70 °C; **b**. Boc_2O, DCM, rt, 34% (over 2 steps); **c.** Sonogashira: TMS-acetylene, $Pd(PPh_3)_4$, CuI, Et_3N, Tol, rt, quant.; **d.** TBAF, THF, rt, quant.; **e.** $KMnO_4$, $NaHCO_3$, tBuOH/H_2O, 30 °C, 50% (over 2 steps).

R = H R' = H	R = I R' = H	R = I R' = Boc	R = C≡C-TMS R' = Boc	R = C≡CH R' = Boc	R = COOH R' = Boc
N-methylbenzylamine	**61** -	**13** 34%[a]	**62** quant.	**30** quant.	**34** 50%[a]
KL08	**63** -	**60** 64%[a]	**64** 76%	**65** 90%	**66** 67%

Table 5.4. KL07 and **KL08** derivatives. Reagents and conditions as shown in **Scheme 5.7**. [a]Over two steps.

With the modular synthetic framework fully validated, D. Lanaro directed efforts towards *de novo* synthesis of fragments, with a particular focus on **KL07**. Building blocks in **Scheme 5.8** were acquired to be precursors of the iodinated versions of **KL07** and **KL08**, among others.

Scheme 5.8. *De novo* synthesis of fragments: achievements and perspectives*



***Reagents and conditions: a.** $MeNH_2$·EtOH, DCM, rt. **b**. Boc_2O, DCM, rt, 81% (over 2 steps). Y: carbon or heteroatom.

De novo synthesis of fragments opens the door to derivatizing fragment structures in order to perform SAR studies in the long term. For now, *de novo* synthesis of the **KL07** series led to pure para-linked final molecules, proving the worthiness of the synthetic effort. The same remains to be done for other fragments.

Thus, fragment functionalization was successful to a certain extent, with the caveat of needing a better strategy for iodination at the desired positions. *De novo* synthesis of fragment structures seems to be the solution for this problem but slows down the speed granted by the modular framework. This setback is counterbalanced by two strategies that complement the modular synthesis. First, the validated FGI series guarantees completion of the whole synthesis if the iodinated fragment is attained. Second, *de novo* synthesis is underway and employs building blocks that can be used for parallel synthesis of new fragment derivatives. With these strategies, the panel of final molecules can rapidly grow and provide valuable information upon evaluation against the target.

5.5. Proofs of concept towards multivalency

As seen in the introduction, multivalency is a staple of targeting lectins. In the case of BC2L-C-N_{ter}, three binding sites are presented on the top of the barrel-like structure, in such a way that favours multivalent interactions. Furthermore, the new crystal structures show an oligosaccharide binding mode established perpendicularly to the barrel's surface, indicating that the native interactions of the superlectin are probably multivalent. Indeed, early probing at this lectin has already used multivalent fucosides, with promising results: a tetravalent calix[4]arene fucoside is the best synthetic ligand known to date.[151]

On the other hand, before attempting multivalency, optimization of monovalent ligands is worth the effort and beneficial in the long term.[58, 61] For this, keeping multivalency in mind is important during glycomimetic design, to ensure it can be implemented at a later date. In the case of the presented C- and N-fucoside glycomimetics, one logical option would be to implement a multivalent handle in the fragment moiety. Indeed, modification of the fragment is less susceptible to be detrimental to affinity. Nevertheless, a different consideration draws attention to the fucoside moiety for implementing a multivalent handle: the binding mode of the oligosaccharide ligands.

Indeed, the new crystal structures of BC2L-C-N_{ter}/oligosaccharide complexes revealed that the key to achieving micromolar affinity is in the third sugar unit: GlcNAc or GalNAc. More importantly, these units consistently established interactions with key residue **Ser82**. Although this residue and its environment were not taken into consideration for the design of 1st generation ligands, its potential to multiply affinity cannot be ignored. Another discovery from the same study was that the monosaccharide L-galactose presented a similar affinity for BC2L-C when compared to its fucoside counterpart (2.0 *vs.* 2.7 mM).[165] L-galactose is an unnatural sugar related to L-fucose, the only difference between these two is the absence of a hydroxyl group in the position 6 of the latter. Keeping these two ideas in mind, synthetic efforts were made towards proof-of-concept structures which could be advantageous for affinity while accommodating multivalency handles.

In parallel with the main fucoside synthetic project, N. Quadrio was entrusted with the task of performing an equivalent synthesis using L-galactose (L-gal) as starting material. Indeed, one of the advantages of the modular synthesis is the ability to swap the sugar unit for an equivalent monosaccharide. As predicted by *in silico* docking, the primary alcohol exclusive to L-gal is hypothesized to establish additional H-bonding with nearby residues **Thr83** or **Tyr48** or participate in water-mediated H-bond networks (see **Figure 5.6**).

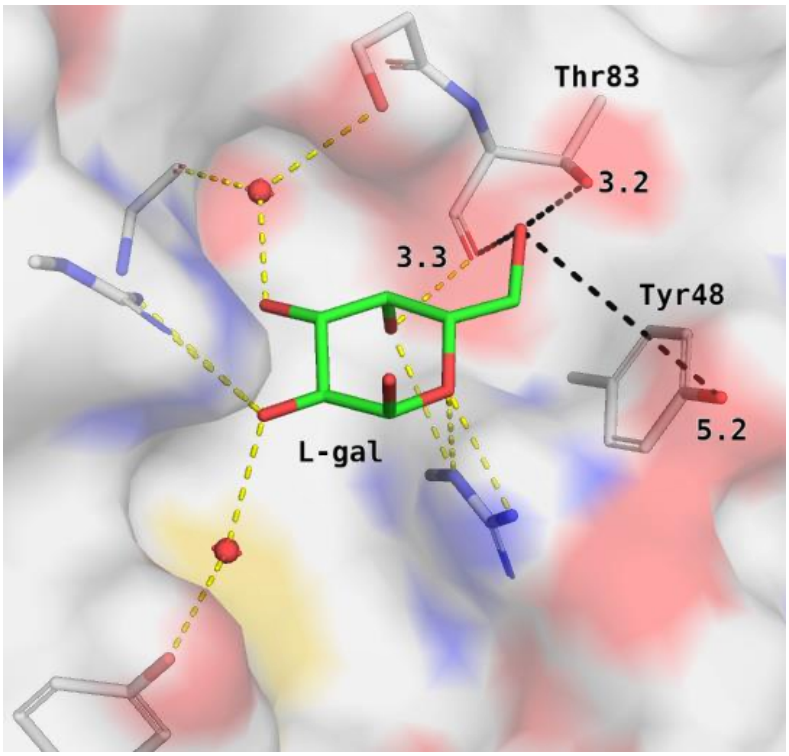
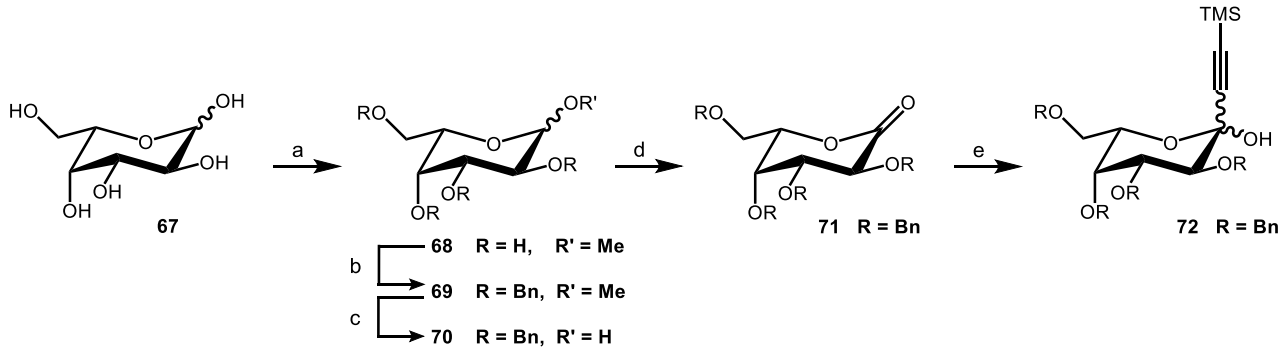


Figure 5.6. Docking of L-galactose in BC2L-C-N_{ter}'s CRD. The known binding mode of fucose is conserved. Distances from O6 to oxygens of residues **Thr83** (3.3, 3.2 Å) and **Tyr48** (5.2 Å) in black.

Furthermore, the fact that L-galactose is not a naturally-occurring sugar gives it a metabolic advantage: bio-machinery designed to metabolize L-fucosides may not be able to recognize and degrade L-galactosides in the same manner. Thus, glycomimetics featuring this unnatural sugar may have an edge as potential glycodrugs.

In the time allotted to this proof of concept, the synthesis towards L-β-galactosylacetylenes proceeded as detailed in **Scheme 5.9**, stopping at molecule **72**, before the β-selective dehydroxylation. The synthesis proceeded smoothly, although with suboptimal yields. The experimental conditions were lightly modified and will need to be optimized if this route is revisited in the future. Since the primary alcohol at C6 has a different reactivity and is relatively solvent-oriented compared to the rest of the sugar, it will be an ideal point to attempt attachment to a multivalent scaffold.

Scheme 5.9 Synthetic route towards L-β-C-galactosylacetylenes*



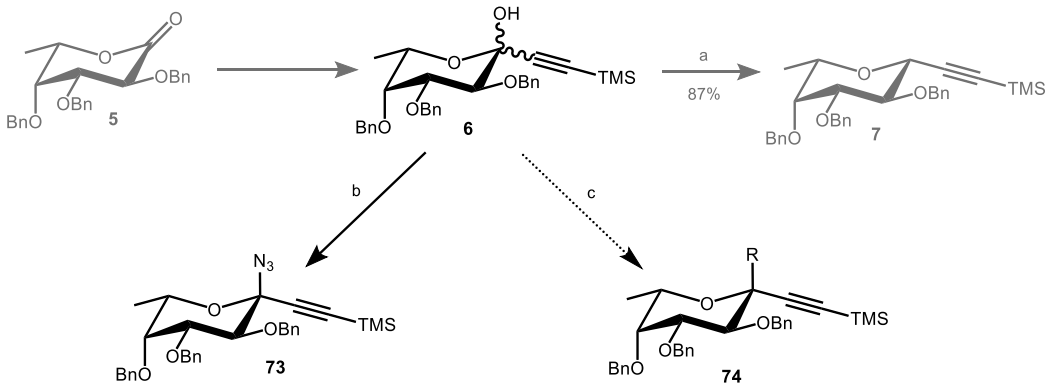
*Reagents and conditions: **a.** Amberlite® IR_{120} H^+, MeOH, 65 °C, 53 %; **b.** BnBr, KOH, Dioxane, 101 °C, 34 %; **c.** HCl, AcOH, 70 °C, 52 %; **d.** I_2, K_2CO_3, DCM, rt, 53 %; **e.** TMS-acetylene, nBuLi, $CeCl_3$, THF, -78 °C, 50 %.

A second interesting approach that can be considered for the sugar moiety is to try to emulate the structure of H-type oligosaccharide ligands. Indeed, these are α-linked fucosides, which directs the reducing end of the oligosaccharide towards the solvent and site Y, where the third sugar unit establishes its contacts. An interesting idea for ligand design would be to target site Y and site X simultaneously by attaching both α- and β-substituents to the fucose core. Such structures have been obtained in the past, by modifying the β-selective dehydroxylation step: the hydride nucleophile can be replaced by other moieties to be selectively α-linked to a β-glycosylacetylene.[203-205]

Clearly, this represents an opportunity for a multivalency handle but also to grow the ligands towards site Y in order to target it by emulating the oligosaccharide interactions. Eventually, both uses could be combined but the approach needs to be validated first, as this method hasn't been attempted on fucosides. Test reactions aiming to install an azido group were

performed using TMS-N_3 and, whereas they ran to completion, the obtained product **73** (**Scheme 5.10**) was unstable, slowly degrading over time. Future probing of this method should be attempted to obtain other substituents, for example by using TMS-CN and allylsilane.

Scheme 5.10 Synthesis of (α,β)-substituted C-fucosylacetylenes: achievements and perspectives*



***Reagents and conditions: a.** Et_3SiH, $BF_3{\cdot}Et_2O$, CH_3CN/DCM, -10 °C, 87 %; **b**. TMS-N_3, $BF_3{\cdot}Et_2O$, CH_3CN/DCM, -20 °C, 47 %; **c. R-TMS**, $BF_3{\cdot}Et_2O$, CH_3CN/DCM.

Thus, two attempts have been made towards growing the designed ligands into new directions, potentially leading to increased affinity and multivalency handles. Although these attempts were cut short due to prioritization, the preliminary proofs of concept didn't show any major obstacle to their completion.

5.6. Outlook

Recapitulating, monovalent fucosides were designed to target BC2L-C-N_{ter}'s CRD and its vicinal site X. This was possible thanks to the collaboration of K. Lal, with computational analysis of the CRD and fragment screening for the vicinal site. The molecules designed featured a fucose and a fragment moiety, β-linked by either alkyne, alkene, amide or triazole functions.

Onwards, a modular synthetic framework towards these molecules was drafted and validated, meeting its bottleneck at the functionalization of fragments. While solutions are being developed to overcome this impasse, a handful of fragments allowed to validate the entirety of the framework. Thus far a panel of β-C- and β-N-fucosides final molecules summarized in **Table 5.5** have been generated, to be screened against BC2L-C-N_{ter}.

Furthermore, side-projects were briefly explored, which have the potential to spring a 2nd generation of BC2L-C-N_{ter} ligands adapted for multivalency.

The synthetic work presented was partly performed by mentored Bachelor and Master students N. Quadrio, D. Ruggeri and D. Lanaro, who were especially helpful by continuing the synthetic efforts after the end of the period at the University of Milan. All things considered, a new framework has been established for straightforward synthesis of a variety of C- and N-fucoside glycomimetics. This easily translatable framework is certainly useful for the long term of this project but can also benefit any other endeavour requiring fucoside 'glycodrugs'.

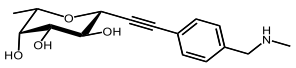
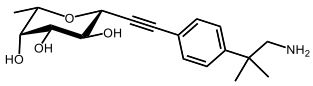
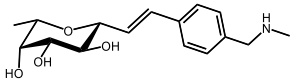
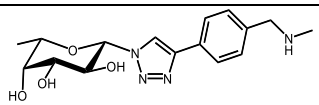
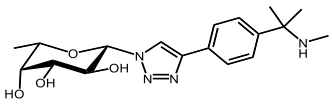
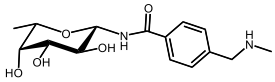
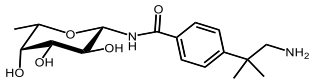
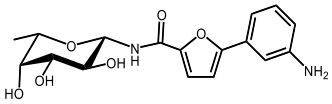
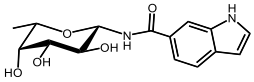
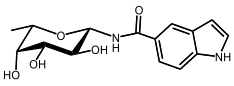
	BC2L-C-N_{ter} antagonist	Molecular weight (Da)	Water soluble
16		**TFA salt: 405.37** (291.35)	**Yes**
27		**TFA salt: 433.42** (319.40)	**Yes**
20		**TFA salt: 407.39** (293.36)	**Yes**
33		**TFA salt: 448.40** (334.38)	**Yes**
40		**TFA salt: 476.45** (362.43)	**Yes**
37		**TFA salt: 424.37** (310.35)	**Yes**
45		**TFA salt: 452.43** (338.40)	**Yes**
47		348.36	**No (DMSO)**
49		306.32	**Yes**
51		306.32	**Yes**

Table 5.5. Panel of final molecules obtained, to be evaluated against BC2L-C-N_{ter}.

6. EVALUATION OF ANTAGONISTS

6.1. Summary

Moving back to CERMAV (University of Grenoble) in March 2020, the final part of the project started, which validated the work performed so far. The evaluation of the newly produced molecules against BC2L-C-N_{ter} was programmed in parallel to the synthetic work continued in the University of Milan by D. Ruggeri and D. Lanaro. This parallel work benefitted from feedback between the evaluation of existing structures and the design and production of new ones.

Prior to this, STD NMR experiments were performed in Milan with the help of Prof. F. Vasile. These experiments were the first to validate the newly synthesized molecules as ligands of BC2L-C-N_{ter} and were continued later on to further characterize the interaction. At CERMAV, two types of biophysical methods were devised to provide a reliable system of ligand evaluation. First, a qualitative method would allow rapid ranking of the ligands while remaining economic in terms of protein and ligand expenditure. The affinity of top-ranked ligands would then be quantitatively assessed by a second method, more costly in terms of materials.

For the first method, fluorescence polarization (FP) and surface plasmon resonance (SPR) were considered, with the latter being preferred. Later on, differential scanning calorimetry (DSC) was also briefly examined. On the other hand, isothermal titration calorimetry (ITC) remained the method of choice for reliable measure of affinity. Although the low affinity of BC2L-C-N_{ter} proved to be a limit in all cases, protocols were established for SPR and ITC, leading to evaluation of all the ligands by at least one technique. Among the results, the alkyne-bound molecules proved to be the strongest binders, with molecule **27** showing an almost 9-fold affinity increase compared to the monosaccharide. Parallel to this, crystallographic efforts led to solving two new crystal structures of protein/ligand complexes featuring BC2L-C-Nter and molecules **27** and **47**. Analysis of these crystal structures further validated the *in silico* groundwork and provided valuable feedback for ligand design.

Thus, three channels were established for the evaluation of antagonists, revealing one 'hit' compound and two crystal structures, opening the way for SAR studies and optimization towards a 'lead' structure aiming towards functional inhibition of BC2L-C-N_{ter}.

6.2. Validation: STD-NMR

The first evaluations of the newly generated molecules was performed through STD-NMR, knowing that experiments had already allowed to validate the binding of αMeFuc to BC2L-C-N_{ter}, as well as binding of fragments **KL03**, **KL08** and **KL09** in presence of αMeFuc.[167] Thus, we intended to expand those results by probing BC2L-C-N_{ter} with final molecule **27**.

In order to evaluate molecule **27** against the lectin, two proton experiments were performed, irradiating aliphatic (-0.05 ppm) or aromatic (10 ppm) residues of the protein, which then transferred their energy to the small molecule in close proximity. The aliphatic irradiation revealed strong signals for the C6 methyl group of the fucoside (1.26 ppm), indicating close contact to the protein, as had been previously observed for the monosaccharide (see **Figure 6.1**).

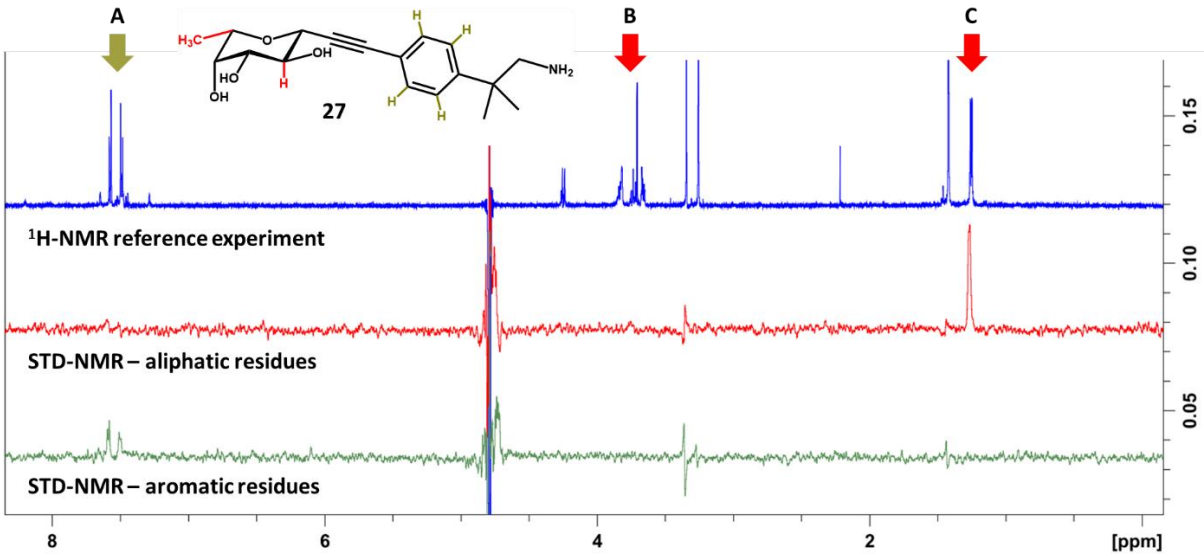


Figure 6.1. STD NMR experiment for **27**. The listed signals correspond to the aromatic protons (A, green, 7.49-7.58 ppm), the C2-bound H2 proton (B, red, 3.74 ppm) and the C6 methyl group protons (C, red, 1.26 ppm).

Weaker signals were also observed, corresponding to H2 from the fucoside ring (3.74 ppm) and the aromatic signals of the fragment moiety (7.49, 7.58 ppm). On the other hand, aromatic irradiation revealed only the aromatic signals, hinting at close contact to aromatic

residues (for example, π/π stacking). By simultaneously observing signals corresponding to the two moieties, the interaction of **27** with BC2LC-N_{ter} was undisputable.

Indeed, these results confirmed the known fucoside binding mode: the main protons exposed to the protein belong to C2 and C6. Combining this to the observed fragment-related signals confirms the expected binding mode of the ligand as designed and later confirmed crystallographically (see section **6.5**). With these results, STD NMR confirmed its potential role as validator of final molecules. To confirm the expected binding, an STD experiment would reveal simultaneously the known sugar signals and new fragment-related signals. By focusing on the latter, the extent of the shape complementarity between fragment and protein site can be assessed. Furthermore, developing these experiments could eventually lead to ranking the molecules and even quantifying the interaction. For now, STD-NMR enabled the rapid validation of new molecules as they were synthesized.

6.3. Qualitative evaluation: FP, SPR, and DSC

a. Fluorescence Polarization

To follow up the encouraging results obtained by STD-NMR, it was relevant to establish a method in CERMAV to assess ligand affinities. Logically, ITC was the method of choice to compare new affinity measures to the existing ITC values of oligosaccharide ligands (see **Table 4.1**). Nevertheless, ITC experiments are costly in terms of protein and ligand, requiring *circa* 0.07 µmol (1 mg) of BC2L-C-N_{ter} and up to 2 µmol of ligand (*ca.* 0.75 mg). Therefore, it was relevant to consider a technique requiring less material, which would provide preliminary affinity assessments or at least allow ranking the fragments for prioritization.

The first technique considered was fluorescence polarization (FP), which had been recently used for probing multivalent fucosides against other fucose-binding lectins. This work had been performed by M. Duca in her work as a member of the PhD4GlycoDrug consortium. As seen in **Figure 6.2**, FP proceeds by measuring the free/bound equilibrium of a fluorescent probe. The fluorescence polarization changes in function of this ratio due to the difference in rotation speeds of the free and complexed molecules. The prototype competition experiment requires constant concentrations of protein and probe and increasing concentrations of

competitor ligand, thus measuring an EC_{50} (half maximal effective concentration). Prior to the competition experiment, it was necessary to define the working concentration of protein [C_{prot}], defined as its EC_{50} against the probe. At this concentration of BC2L-C-N_{ter}, the added antagonists could properly compete against the fluorescent probe. Thus, this concentration would determine the quantity of protein required for each experiment, a key decision factor.

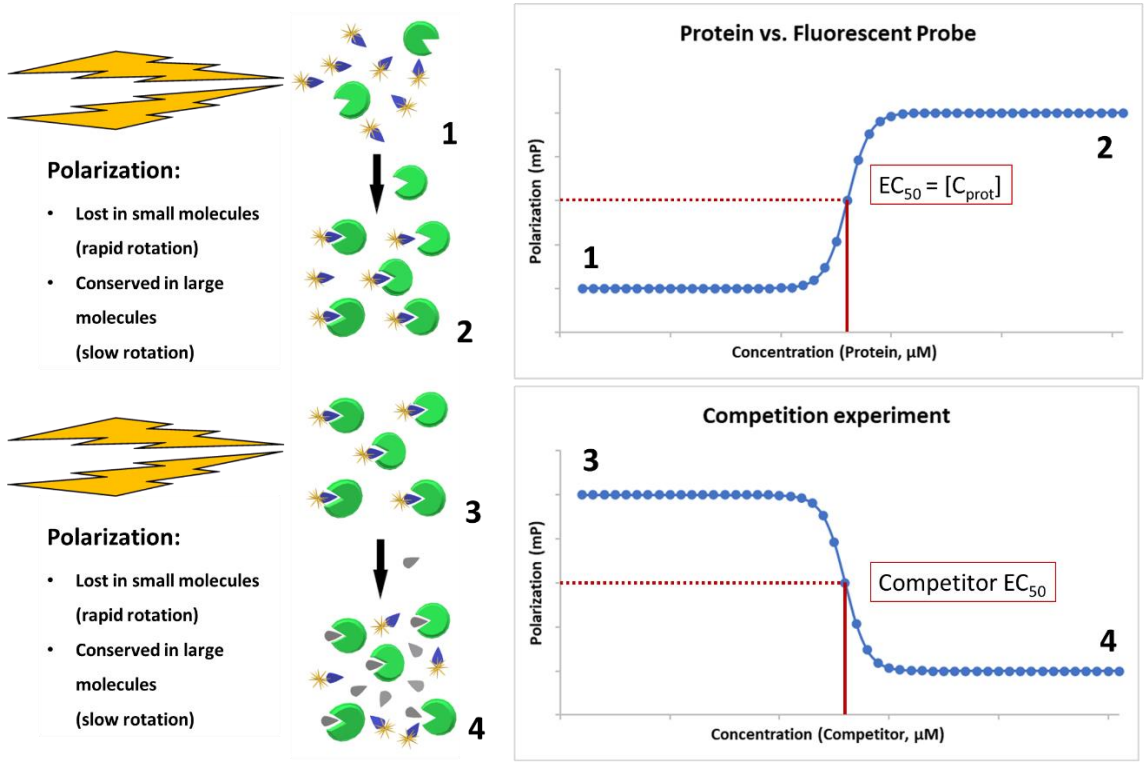


Figure 6.2. Principle of FP. Setup for competition experiments and determination of competitor EC_{50}.

Therefore, increasing amounts of BC2L-C-N_{ter} needed to be added to the minimal detectable concentration of probe (α-L-fuc-FITC: 2nM), which was kindly provided by M. Duca (see **Figure 6.2** and **6.3**). A number of preliminary experiments with increasing concentrations of protein led up to using 1063 µM (14.9 mg/mL) for the most concentrated point of **Figure 6.3**. This last experiment was able to fit the hypothetical sigmoid shape and determine the concentration [C_{prot}] to be used for future competition experiments: 259 µM. Naturally, this value (3.6 mg/mL) was too high to be viable and would require even more protein per experiment than ITC (2 mg).

The need for such concentrated protein derives from the low affinity of BC2L-C-N_{ter} for monosaccharidic fucoside ligands, as seen in **Table 4.3**. In the low millimolar range, this interaction doesn't fit the classic affinities measured in FP competition experiments.[128] Thus, performing FP with a low-affinity fluorescent probe wasn't viable. Although generating a

micromolar fluorescent probe with an H-type epitope could be a solution to this limit, it wasn't attempted. Instead, FP was side-lined to give way to SPR experimentation.

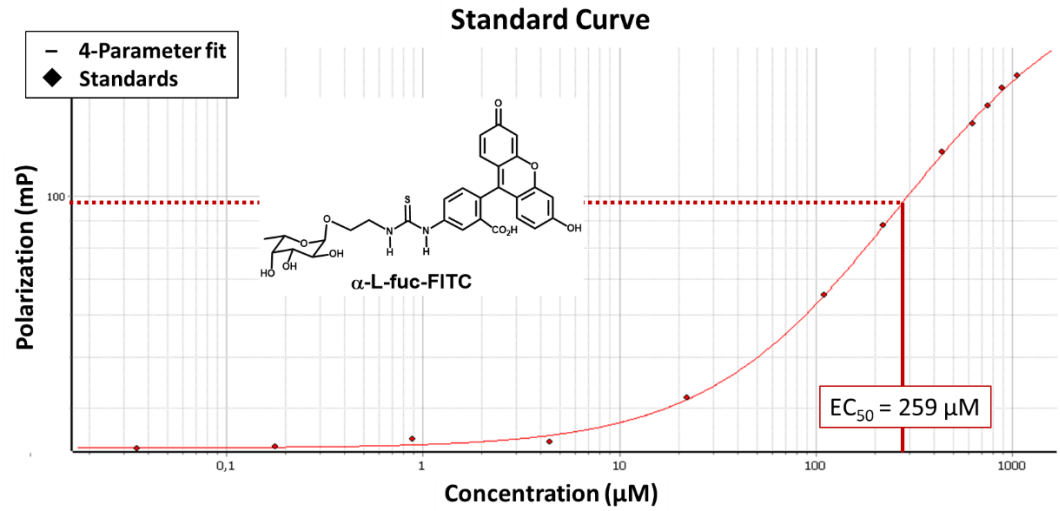


Figure 6.3. Fluorescence Polarization experiment to determine the EC_{50} of BC2L-C-N_{ter} against α-L-fuc-FITC (2nM). The half maximal effective concentration was approximated to [C_{prot}] = 259 µM.

b. Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a surface-based method in which one partner of the interaction is bound to a chip while the other partner flows above it. When a chip featuring one of the partners isn't available, competition experiments can be an alternative. Thus, competition experiments were planned with the existing chip exhibiting tethered monosaccharidic fucosides. As schematized in **Figure 6.4**, flowing protein at a given concentration and antagonists at increasing concentrations allows to measure the inhibition of the protein/chip interaction. Accordingly, a half maximal inhibitory concentration (IC_{50}) could be obtained for the new antagonists. The first step towards such experiments was to define the working concentration [C_{prot}] at which antagonists could compete against the functionalized surface. Again, this value would be a key decision factor.

To define [C_{prot}], the fucoside chip was assayed against different concentrations of BC2L-C-N_{ter}, ranging [14 - 71 µM] (0.2 – 1.0 mg/mL) as seen in **Figure 6.4**. The expected interaction between protein and surface took place and, once more, it was apparent that the concentrations used were too low to accurately assess [C_{prot}]. Indeed, the concentration recommended at this stage is 2 x K_D.[160] However, such a concentration was too high for an 'economic' protocol and could lead to protein aggregation. Thus, the weak affinity of this interaction became a limit to experimental design once more. Two solutions were considered

to circumvent this issue: to implement either H-type 1 epitopes or BC2L-C-N_{ter} on the surface of new chips. Although the first would likely allow reliable competition measurements, it was the latter that was developed. Indeed, a protein-functionalized chip would allow direct measure of the protein/antagonist interaction affinity instead of indirect IC_{50} values.

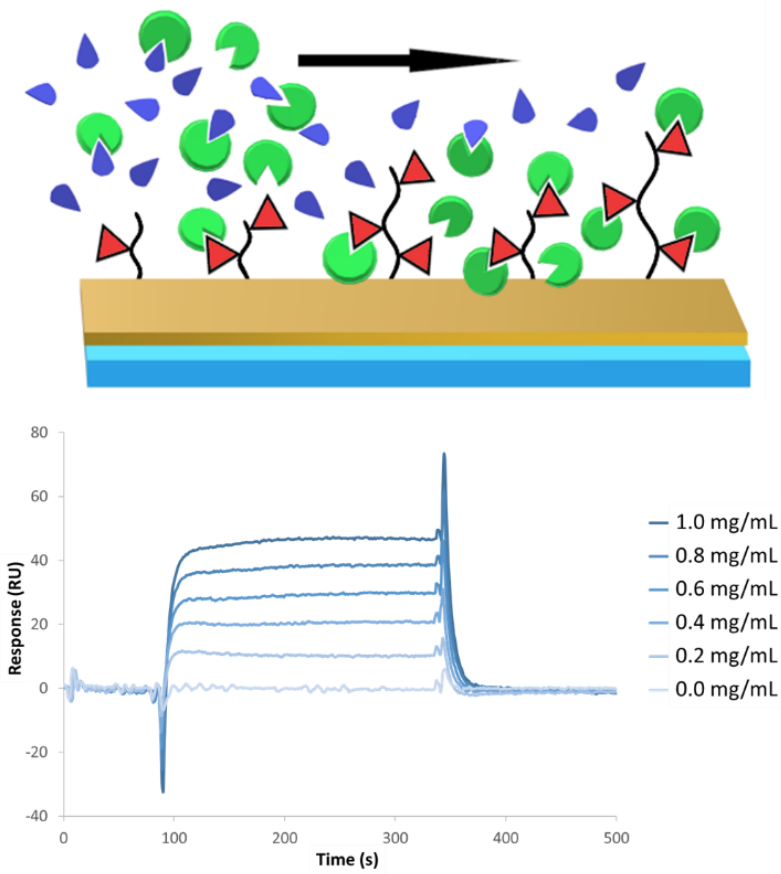


Figure 6.4. Top: Schematic representation of an SPR competition experiment. **Bottom:** BC2L-C-N_{ter} [14 - 71 µM] interacts with the fucoside chip in a dose-dependent manner.

However, before fabricating a protein-coated chip, some competition experiments were attempted with the fucose chip to assess their feasibility at an unoptimized protein concentration. Thus, [C_{prot}] = 43 µM (0.6 mg/mL) was selected, being high enough to provide an observable R_{eq} of *ca.* 30 - 40 RU but remained material-efficient, requiring less than 0.5 mg of BC2L-C-N_{ter} per experiment. Thus, ligands **16** and **27** were diluted to the concentration range [0.1 - 2560 µM] and used in competition against the fucosylated chip, as seen in **Figure 6.5**. These experiments provided a loose approximation of the IC_{50} values, limited by the high concentrations of ligand necessary to complete the sigmoid.

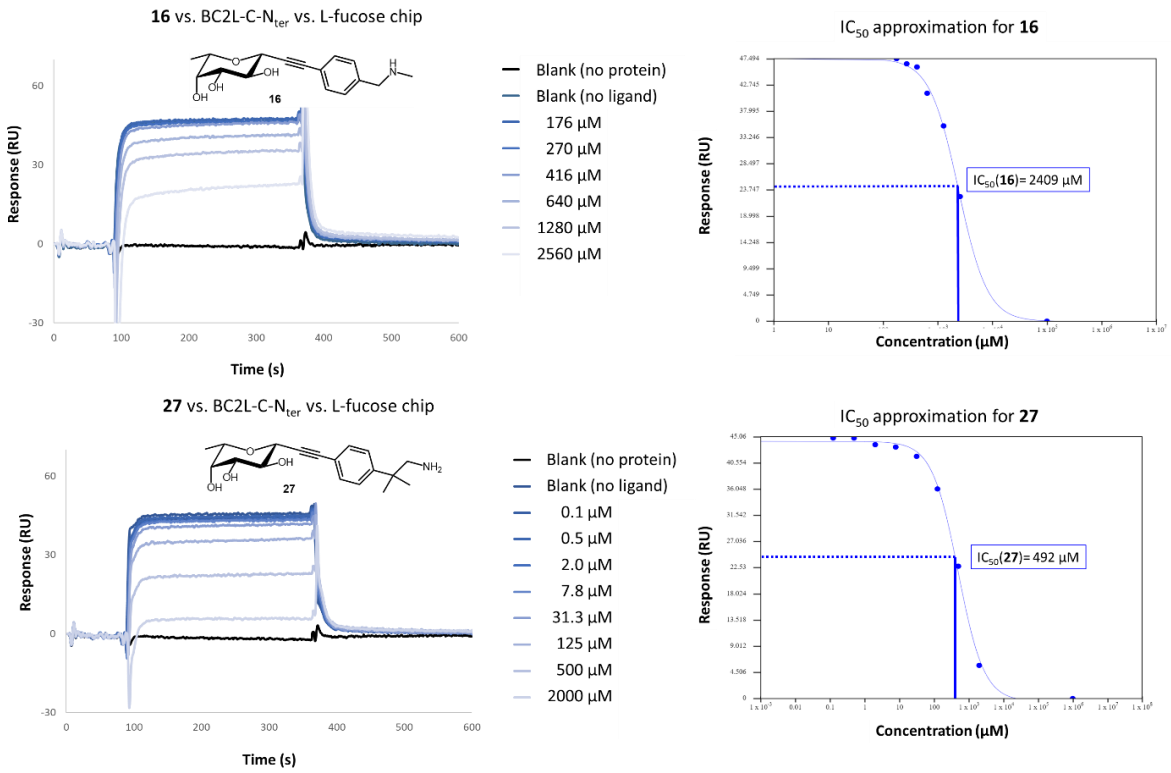


Figure 6.5. SPR experiments: competition of ligands **16** and **27** against the protein/fucose chip interaction. IC_{50} approximation provides a means to classify ligands: IC_{50}(**16**) = 2409 µM and IC_{50}(**27**) = 492 µM.

Indeed, the high concentrations (over 2 mM) needed to attain full inhibition of the interaction were a second limitation for this type of assay, showing that the affinity of the newly synthesized ligands remained in the millimolar range. This is not unexpected, since the ligands are heavily based on the monosaccharide and do not try to mimic the stronger binding observed for oligosaccharides. As similar projects have shown, the competition experimental setup provides reliable data when the affinities measured are in the low micromolar range.[206] Nevertheless, the IC_{50} value obtained in this case could be used as a guide for ranking the synthetic ligands: as it transpires, **27** (IC_{50} = 492 µM) is a better ligand than **16** (2409 µM). Measured in a similar way, the IC_{50} for L-fucose is 543 µM, which would indicate that **16** may not be a ligand or has an abnormal behaviour in this assay.

A second way to rank ligands using competition was tested: a single experiment injecting different ligands sequentially at a constant concentration. **Figure 6.6** showcases a proof of concept for this method, proving it can be a rather fast way to rank and prioritize ligands but does not produce reliable quantifiable data and is only applicable to structurally comparable ligands.

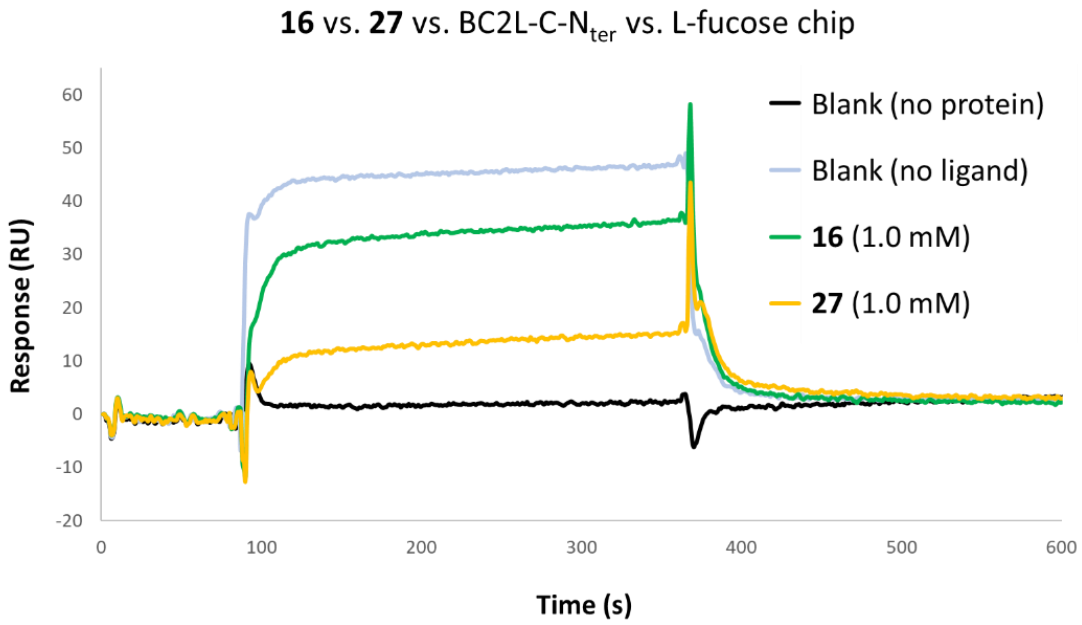


Figure 6.6. Proof of concept for SPR competition experiment: rapid evaluation of ligands at a constant concentration. Ligands closer to the baseline are predicted to have higher affinity.

Pressing forward, a chip was functionalized with BC2L-C-N_{ter}, requiring 7.1 nmol (0.1 mg) of purified lectin, which was in line the aim of reducing protein expenditure. It is worth to point out that the new construct rBC2L-CN2 had demonstrated outstanding stability while in storage. This led to the hypothesis that chips functionalized with this construct would have a longer shelf life compared to other similar lectin chips, which are known to lose binding activity in the range of days to weeks. Thus, the first-ever BC2L-C-N_{ter}-functionalized SPR chip was created with a density leading to a calculated R_{max} = 173 RU for ligands weighing *ca.* 300 Da. It showed the expected affinity for fucosides and extended shelf life (active after 8 weeks when stored in buffer).

As a first test, the activity of the chip was evaluated with known ligands of the lectin: αMeFuc, Fucα(1-2)Gal, H-type and Lewis oligosaccharides as detailed in **Figure 6.7**. This was done in order to establish references for chip performance, which could be re-assessed later to characterize shelf life. Synthetic antagonist **27** was added to this experiment to examine and compare the sensogram profiles of different types of ligand. It was clear that the two ligand families weren't comparable. Firstly, the profile of the synthetic ligand evoked a stronger response than the oligosaccharides and dissociated from the chip with difficulty, which can be attributed to its enhanced hydrophobicity. Secondly, the oligosaccharides couldn't be compared to **27**, or each other, since responses were influenced by the difference in molecular weight: higher R_{eq} and R_{max} are expected for heavier ligands, which explains the

ordering observed. Regarding the lighter saccharides, it was unsurprising to see low responses for the mono and disaccharide, since there is a leap in affinity between these and the rest.

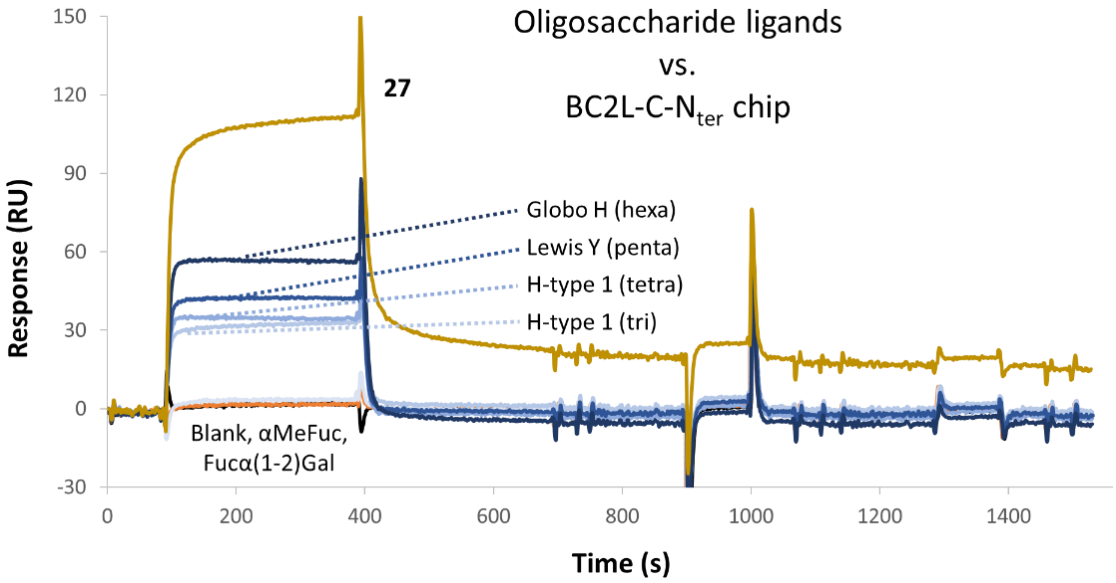


Figure 6.7. Evaluation of the new BC2L-C-N_{ter} SPR chip against blood-type oligosaccharide ligands and **27**. All ligands concentrated at 300 µM. Regeneration steps at 900 (5mM fucose) and 1100 seconds (buffer).

Onwards, it became essential to find proper conditions to regenerate the surface and return to the baseline after injection of synthetic ligands, in particular **16**, which was systematically difficult to clear from the chip, hinting again at an effect unrelated to lectin affinity. The regeneration conditions seen in **Figure 6.7** were eventually optimized and implemented as necessary, sometimes sequentially after high-concentration injections.

With this preliminary work established, it was possible to define an experimental procedure for affinity assessments. Thus, the synthetic ligands were dissolved and sequentially injected in concentrations ranging [3.3 – 3500 µM]. Once again, it was apparent that high concentrations of ligand were needed to better assess affinity (see **Figure 6.8**), but the ligand concentration was capped at 3.5 mM for the benefit of later experimentation. The experimental procedure allowed the assessment of affinity through the steady state, which was easy to attain. On the other hand, the sensograms obtained couldn't be fitted for kinetic evaluation.

Table 6.1 summarizes the results obtained for duplicates of each experiment. Experiments performed shortly after each other were consistent but it was noticed that measurements taken weeks apart led to drifting values for the same ligand. For example, experiments with **51** were performed the same day, with low standard deviation (**Figure 6.8**, top). On the other

hand, the first measure of **47** (**Figure 6.8**, bottom) returned K_D = 1.67 mM, whereas the second measure, performed 52 days later, had a similar profile but returned K_D = 3.05 mM.

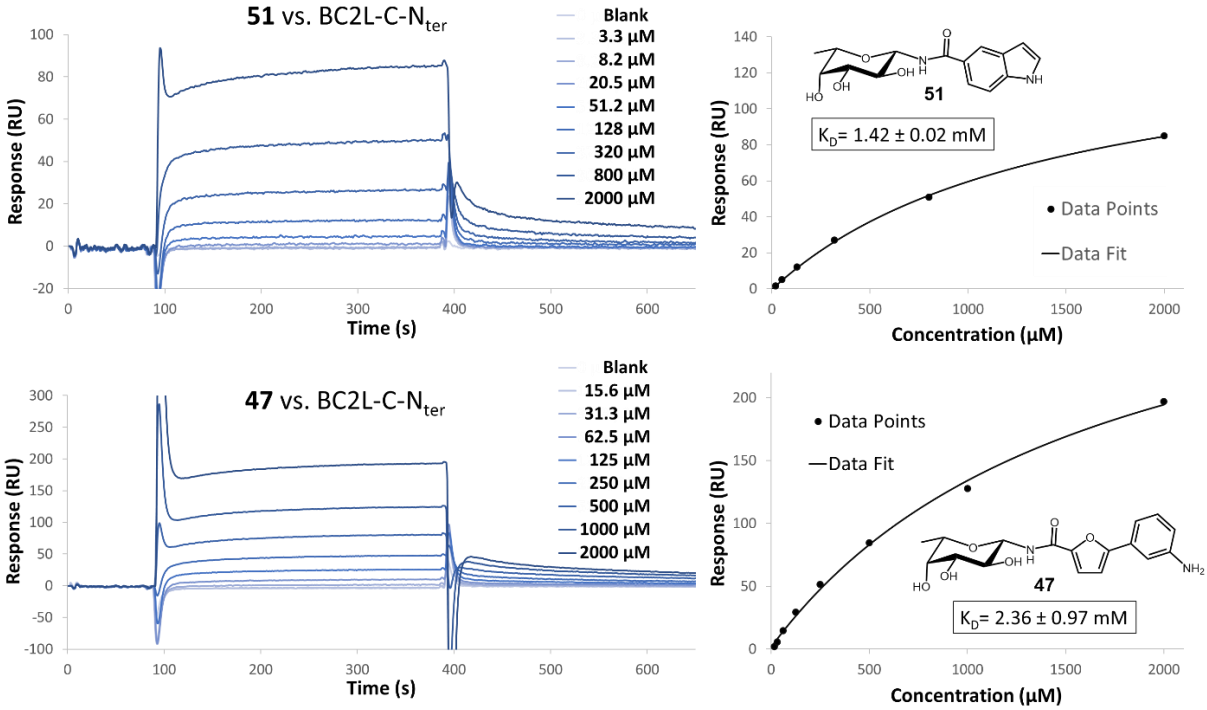


Figure 6.8. Examples of definitive SPR experiments. One representative example of duplicates. K_D: standard deviations from two experiments.

This is not unexpected: the chip's activity gradually degrades over time, as is usual for protein chips. Another factor leading to variability was the addition of 8% DMSO to solubilize otherwise insoluble ligand **47**: although all buffers used in the experiment were adjusted to 8% DMSO, the ensuing sensograms showed heightened buffer effects and buffer mismatch peaks (**Figure 6.8**, bottom). It follows that, in order to make the most of the SPR affinity measures, it's preferable to perform measures in a short period of time, when enough new structures have accumulated. Furthermore, results from experiments requiring DMSO have to be adjusted to DMSO-free (normal) experiments, for example by measuring the same ligand in normal and DMSO conditions and using its affinity as a reference for the rest of the DMSO-dissolved ligands.

	BC2L-C-N_{ter} antagonist	SPR affinity [mM]	ITC affinity [mM]
16		7.85 ± 3.39	1.24 ± 0.07
27		1.33 ± 0.15	**0.28 ± 0.01**
20		1.02 ± 0.02	3.37 ± 0.40
33		2.45 ± 0.02	6.25 ± 0.72
40		1.19 ± 0.05	2.49 ± 0.06
37		1.57 ± 0.06	3.66 ± 0.21
45		0.94 ± 0.01	2.55 ± 1.00
47		2.36 ± 0.97	-
49		3.42 ± 0.22	3.49 ± 1.30
51		1.42 ± 0.02	-

Table 6.1. SPR and ITC affinity measurements for the panel of antagonists. Standard deviations from duplicates.

As seen in **Table 6.1**, SPR evaluation of the ligand panel resulted in low millimolar affinities, comparable with that of αMeFuc (ITC: 2.43 mM). Excepting molecule **16**, K_D values in the range of [0.9 - 3.4 mM] indicate that all ligands can be considered equivalent, especially since these are low-affinity interactions. Molecule **16** proved to be an outlier, likely due to unspecific interactions with the chip surface since ITC confirmed its affinity to be millimolar and comparable to the other structures. Indeed, retrospective examination using ITC values and crystallographic data have allowed re-examination of the accuracy of SPR data.

For example, ligands **20**, **40** and **45** were measured with a new SPR chip and had promising affinity values around 1 mM. However, this proved to be an artifact of the new chip since ITC data didn't follow the trend of increased affinity. The chip in question was less densely populated with lectin: its calculated R_{max} was of 105 RU, as opposed to the previous R_{max} = 173 RU (for ligands weighing *ca.* 300 Da). This means that the fabrication of the chip has to

be meticulous in order to ensure comparability between measures. Alternatively, work with a new chip can start with a 'standard' measure of a previously evaluated ligand, to establish a comparison point (exactly as mentioned earlier for experiments involving DMSO).

Putting **20**, **40** and **45** aside, the best-performing ligands in SPR were **27**, **51**, **37**, and **47** (first measure, read above), in that order. Of these, **27** and **47** were successfully co-crystallized with the protein, and **27** had sub-millimolar affinity on ITC. Conversely, **51** and **47** could not be measured by ITC due to lack of material but their SPR values indicate the potential worthiness of re-synthesizing the structures to complete ITC evaluation. Finally, molecules **33** and **37** performed better on SPR than on ITC, showcasing the discrepancies often observed between these two techniques (discussed in the next section). In such examples, the affinity of the ligands in question could be measured by a third technique, providing insight on whether SPR or ITC should be trusted in the case of discrepancy. Naturally, it has to be kept in mind that for the low millimolar affinities recorded, these values can be considered to be in the same range. An interesting observation on the retrospective analysis of SPR results: although molecule **16** was an outlier throughout the SPR campaign, such behaviour shouldn't be considered eliminatory but rather interesting to probe by a second technique. Indeed, molecule **16** was the only one besides **27** to show better affinity than the monosaccharide by ITC (K_D = 1.24 mM).

As a final note on this SPR campaign and its limits, two additional experiments were performed. These were designed to evaluate the protocol when applied to ligands of other types and with higher affinities for the lectin. As seen in **Figure 6.9**, one of such ligands was the Globo H (H-type 3) hexasaccharide and the other was a trivalent α-fucoside compound synthesized by M. Duca in the scope of her PhD4GlycoDrug thesis: 'Design and synthesis of multivalent carbohydrate inhibitors of lectin virulence factors'. These two experiments allowed to accurately measure the affinities of the two ligands since the concentrations used remained in the advised range [0.1 x K_D - 10 x K_D]. Indeed, the value obtained for the hexasaccharide was comparable to values obtained by ITC: 32 µM vs 26 µM (**Table 4.3**). On the other hand, the trivalent compound exhibited an avidity of 192 µM, which denotes a clear multivalent effect when compared to the monosaccharide unit (2.43 mM on ITC). Thus, this experiment also highlighted the potential of combining the glycomimetic strategy with multivalency in the future. It is worth to note that the experiments with the trivalent compound were performed on a 61-day old protein chip, showing that although the chip

slowly degrades over time, it has a rather long shelf life. Also important to note: the two experiments measuring the affinity of Globo H were performed on two different chips of different protein density (R_{max} = 173 RU *vs* 105 RU), yet measured to the same affinity accurately: 33.6 µM and 30.9 µM, respectively. This indicates that the variability observed for chip to chip is especially detrimental for measures of low-affinity ligands compared to stronger affinities.

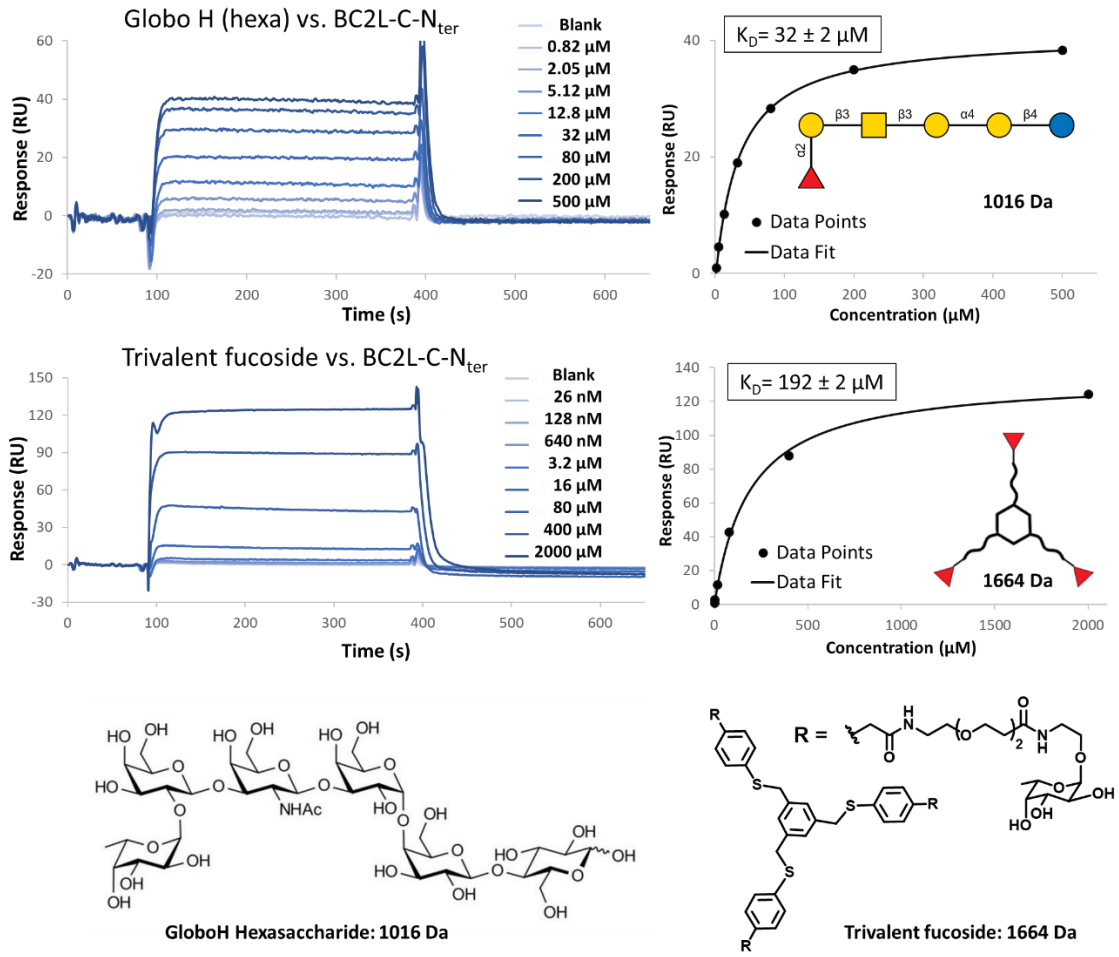


Figure 6.9. Additional SPR experiments. One of two duplicates shown.

Thus, SPR was validated as a useful method to perform material-economic early assessment of BC2L-C-N_{ter} inhibitors but quickly met a first limit in the difficulty of evaluating low-affinity interactions and a second one in the need for speedy measurement campaigns. Nevertheless, SPR measured good affinities for two molecules that would go on to show important results in other avenues: **27** and **47**. Thus, SPR predicts (to a certain degree) the performance of the ligands in a material-economic protocol as it was originally intended. Moreover, low-yielding

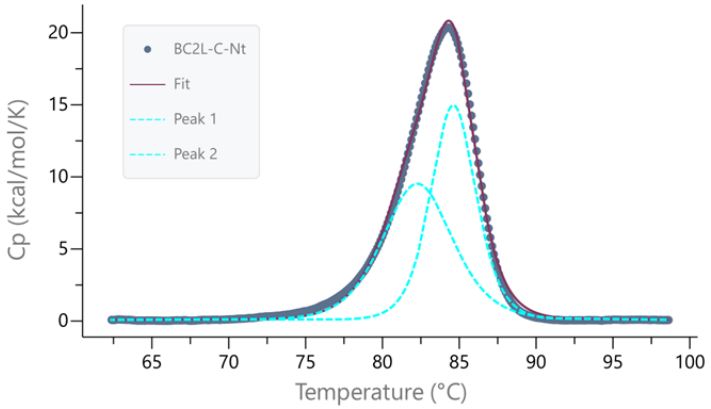
molecules of particularly difficult synthesis can only be assayed by this technique. This was the case for **47** and **51**, whose performance on SPR underlines the interest of an eventual re-synthesis. It remains to be seen whether SPR will become increasingly useful for a 2nd generation of synthetic ligands with improved affinities. For now, SPR screening is a ligand-economic technique for early evaluation of new BC2L-C-N_{ter} ligands and its output should continue to be scrutinised in parallel with other techniques.

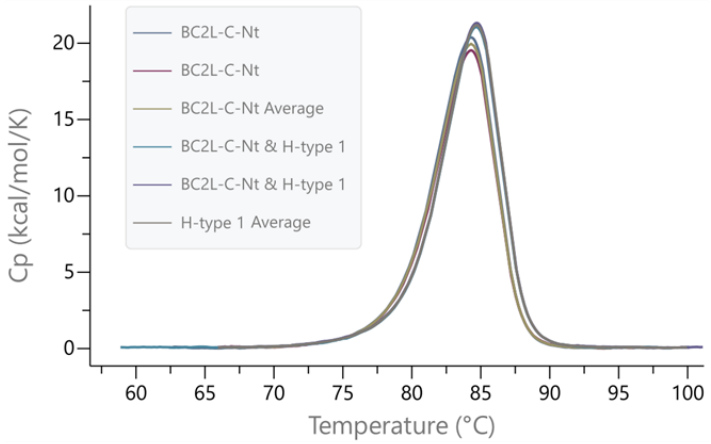
c. Differential Scanning Calorimetry

DSC is a calorimetry technique that was briefly studied in order to assess its potential to complement SPR as a material-economic ligand 'screening' platform. A single set of experiments was performed to evaluate the influence of ligands in the thermal stability of BC2L-C-N_{ter}: the chosen ligands were H-type 1 (trisaccharide) and **27**, which had shown the best affinity in their ligand category. Each experiment required as little 3.6 nmol (0.05 mg) of BC2L-C-N_{ter} and 0.01 µmol of ligand.

As seen in **Figure 6.10**, the construct rBC2L-CN2 is as stable as previous constructs, having a melting temperature of *ca.* 85 °C, possibly the highest measured for this lectin domain.[135] Curiously, the denaturation of the protein was better fitted as two events (Tm_1 = 82.2 °C and Tm_2 = 84.5 °C), which may be attributed to the separation of protomers prior to full denaturation. When ligands were added to the protein in a 1:10 ratio, the shifting of the melting temperatures was measured. The best ligand known for BC2L-C-N_{ter} (H-type 1, K_D = 25µM) stabilized the protein, with a positive shift ΔTm = + 0.4 °C for each of the events. On the other hand, synthetic ligand **27** (K_D = 281 µM) stabilized the protein to a lesser degree with ΔTm = + 0.2 °C for both events. These results confirm no detrimental effect to the stability of BC2L-C-N_{ter} by either type of ligand.

Taking into account the results and the low protein/ligand costs for DSC experiments, less than a tenth compared to ITC, it is conceivable to use this technique for screening pools of potential BC2L-C-N_{ter} ligands. Nevertheless, DSC doesn't measure affinity, so it still needs to be related to another technique in order to confirm any trends observed. On the other hand, tailored protocols of DSC could allow to assess the thermal stability of the new synthetic ligands at the same time as they are screened for the binding interaction.





Ligand	T_{onset} (°C)	Tm_1 (°C)	Tm_2 (°C)	ΔTm (°C)
No ligand	73.94 ± 0.05	82.21 ± 0.03	84.54 ± 0.01	-
27	73.29 ± 1.54	82.43 ± 0.02	84.71 ± 0.01	**(+) 0.17 ± 0.01**
H-type 1	73.94 ± 0.26	82.65 ± 0.01	84.92 ± 0.01	**(+) 0.38 ± 0.01**

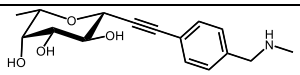
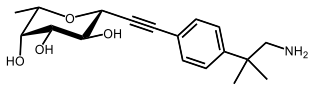
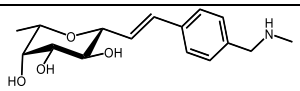
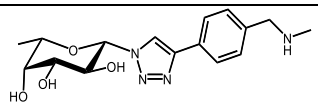
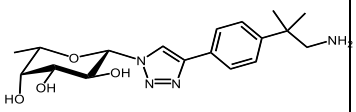
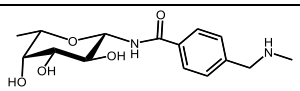
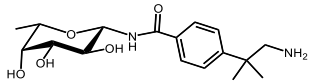
Figure 6.10. DSC experiments **Top:** Fitting with two denaturation events, characterized by melting temperatures Tm_1 and Tm_2. **Bottom:** Representative experiment comparing presence and absence of ligand H-type 1. ΔTm represent the temperature difference between main thermal events (Tm_2). Temperatures: standard deviations from two experiments.

6.4. Quantitative evaluation: ITC

As previously outlined, the evaluation of BC2L-C-N_{ter} antagonists was conceived as a two-step process: first, a 'prioritization' phase, then a material-costly method to quantify the affinity of for the promising structures (or the ones obtained in large quantities). ITC was set to be this second method for two reasons. The first is related to the self-consistency of affinity measures: this technique has consistently been used to rank the oligosaccharide ligands of the lectin, so it should be used to compare synthetic ligands to the previously measured values. The second reason relates to the validity of the affinities obtained: as opposed to other available techniques ITC measures affinity directly in solution and does not require the

modification or tethering of the interacting molecules or any type of competition. Furthermore, ITC allows to assess the thermodynamic parameters of binding, which is essential to characterize the interaction of new ligands with their binding site and improve their performance. However, extensive thermodynamic study through ITC requires affinities in the micromolar range or lower, which was not the case for the current project. Nevertheless, ITC evaluation of low-affinity interactions could still be achieved via 'low c-value' experiments (see Part 2: **RESEARCH METHODOLOGY - Isothermal Titration Calorimetry**).

Indeed, low c-value experiments account for the difficulty of saturating the protein with low-affinity ligands and require ligand concentrations as high as 100 x K_D. Naturally, this leads to large consumption of materials as detailed earlier: 0.7 mg of BC2L-C-N_{ter} and 0.75mg of ligand were consumed in a typical low c-value experiment. In order to ensure saturation of the sites, the working concentration of ligand was brought up from 20 mM to 40-50 mM (*ca.* 20 x K_D), which was the limit for solubility for some of the compounds. Thus, the ITC experiments for antagonists were designed with [C_{prot}] = 200 µM (2.8 mg/mL) and [C_{lig}] = 40-50 mM. For an average 3 mM affinity (K_a = 333 M^{-1}), the c-values of the experiments were c = 0.07. Thus, by fixing the stoichiometry of the interaction to N = 1, affinity values were obtained for the antagonists as seen in **Table 6.1**, replicated below.

	BC2L-C-N_{ter} antagonist	SPR affinity [mM]	ITC affinity [mM]
16		7.85 ± 3.39	1.24 ± 0.07
27		1.33 ± 0.15	**0.28 ± 0.01**
20		1.02 ± 0.02	3.37 ± 0.40
33		2.45 ± 0.02	6.25 ± 0.72
40		1.19 ± 0.05	2.49 ± 0.06
37		1.57 ± 0.06	3.66 ± 0.21
45		0.94 ± 0.01	2.55 ± 1.00

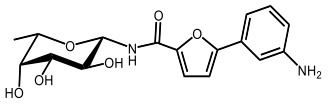
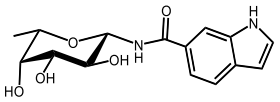
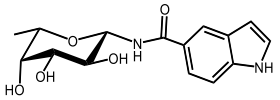
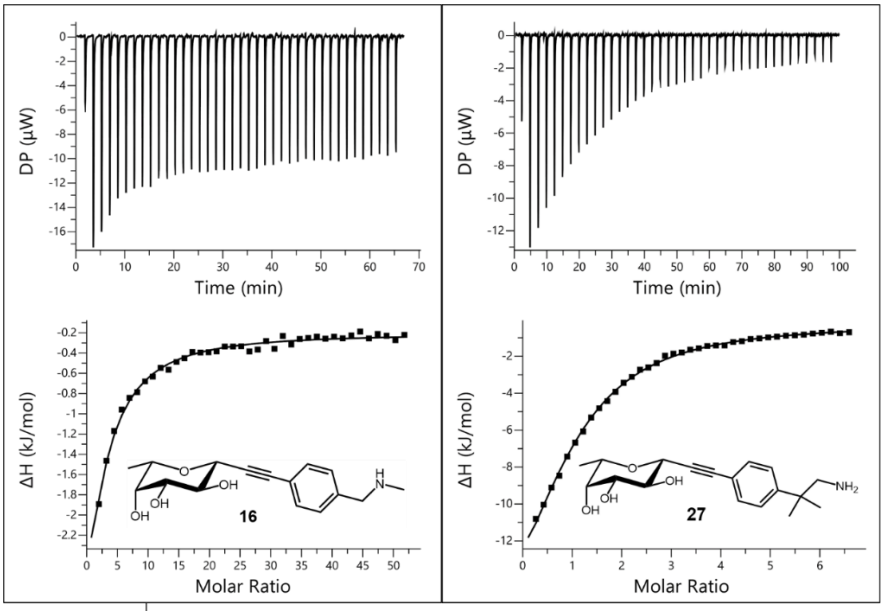
47		2.36 ± 0.97	-
49		3.42 ± 0.22	3.49 ± 1.30
51		1.42 ± 0.02	-

Table 6.1. (bis) SPR and ITC affinity measurements for the panel of antagonists. Standard deviations from duplicates.

Thus, the K_D values obtained were spread in a wider range than SPR: [281 – 6250 µM], allowing a better sense of which structures performed worse than the original monosaccharide, and which could be 'hits'. Of the evaluated panel, **16** and **27** surpassed the monosaccharide with affinities of 1240 and 281 µM, respectively. Although it showed abnormal behaviour on SPR, **16** showed a two-fold increase of affinity from αMeFuc (2.43 mM). On the other hand, **27** cemented its role as best-performing ligand in both techniques, with a nearly 9-fold affinity increase. The successful increase of affinity upon adding alkyne-bound fragments to L-fucose validated both this type of linker and, more importantly, the fragment screening and ligand design strategies.



Ligand	N (fixed)	K_D (µM)	−ΔG (kJ/mol)	−ΔH (kJ/mol)	−TΔS (kJ/mol)
16	1	1237 ± 70	-16.6 ± 0.2	-16.6 ± 0.5	0.0 ± 0.3
27	1	281 ± 11	-20.2 ± 0.2	-21.4 ± 0.8	1.2 ± 0.6

Figure 6.11. Examples of low c-value ITC experiments. One representative example of triplicates. Standard deviations from two or more experiments.

Although thermodynamic values obtained in low c-value experiments aren't reliable and should be taken with reservation, an interesting trend was observed exclusively for ligands **16** and **27**. The entropic component contributing to their binding free energy was virtually non-existent (see **Figure 6.11**). This low entropy component was hypothesized when the ligands were designed: a strictly rigid linker would force the bio-active binding conformation at all times, minimizing the entropic costs for the binding event.

Thus, low c-value experiments allowed to accurately evaluate the panel of glycomimetics against BC2L-C-N_{ter}, despite the low affinity of their interactions. Most of the affinities measured hovered around the original affinity of the monosaccharide, and in the millimolar range can be considered equivalent. Nevertheless, the two alkyne-bound ligands **16** and **27** performed particularly well and showed an interestingly low entropic value. Although thermodynamic assessments cannot be trusted until higher affinities are reached, the measured 2-fold and 9-fold affinity increases are an important first step in that direction. It remains to be seen if higher-affinity alkyne ligands will continue to show this thermodynamic profile. Another point to be seen is the eventual evaluation of structures **47** and **51**, which performed well in SPR but couldn't be evaluated on ITC due to lack of material. In the meantime, the crystal structures can be used to assess the interaction of **47** and further rationalize the affinity observed for **27**.

6.5. Crystallography

Two crystal structures featuring complexes of BC2L-C-N_{ter} with synthetic ligands **27** and **47** were obtained, allowing ground-breaking structural study of the interaction of this lectin with synthetic ligand. The structures resulted from BC2L-C-N_{ter} crystals pre-grown in the usual condition: 1.2 M sodium citrate at pH 7.0, which were soaked for over 24 h in the same growing condition, but in presence of high ligand concentration. The crystals were cryo-protected using 2.5 M sodium malonate at pH 5 and diffracted to high resolution. The resulting structures were solved to 1.79 Å and 1.32 Å by molecular replacement. The relevant statistics can be found in **Table 6.2**.

Table 6.2 Data collection and refinement statistics. Values in parentheses are for highest-resolution shell.

Data Collection	**BC2L-C-N_{ter}/27**	**BC2L-C-N_{ter}/47**
Beamline	Proxima 2A (Soleil)	Proxima 1 (Soleil)
Wavelength	0.98011	0.97857
Space group	P63	P63
Unit cell dimensions (Å,°)	a = b = 44.0, c = 94.1	a = b = 42.9, c = 94.9
Resolution (Å)	47.07-1.79 (1.83-1.79)	47.44-1.32 (1.34-1.32)
Nb/nb unique reflections	196,154/9,696	470,817/23,283
R_{merge}	0.039 (0.228)	0.054 (0.480)
R_{meas}	0.040 (0.244)	0.057 (0.504)
Mean I/σI	49.1 (11.1)	30.6 (6.7)
Completeness (%)	99.70 (95.4)	100.0 (100.0)
Redundancy	20.2 (17.2)	20.2 (20.1)
CC 1/2	1.000 (0.990)	1.000 (0.959)
Refinement (in progress)		
Resolution (Å)	38.07-1.79	37.18-1.32
Nb/nb free. reflections	9,670/454	23,246/1,163
R_{work}/R_{free}	13.7/19.4	10.5/13.3
Rmsd Bond lengths (Å)	0.016	0.013
Rmsd Bond angles (°)	1.93	1.68
Rmsd Chiral (Å^3)	0.089	0.081
No. atoms/Bfac (Å^2):		
Protein	983/27.6	1,051/14.8
Ligand	23/30.0	25/13.2
Waters	100/35.8	137/26.0
Ramachandran Allowed (%)	100	100
Favored (%)	97.7	97.4
Outliers (%)	0	0

Upon initial inspection of the electron density, it was clear that the fucose-binding site was occupied rather than filled with water molecules, as seen for apo crystals of BC2L-C-N_{ter}. Instead, density for a fucoside and a β-anomeric-oriented substituent were apparent (see **Figure 6.13**), confirming the intended binding mode observed when the molecules were docked by K. Lal (see **Figure 6.12**). In the case of **27**, the fragment moiety pushed **Tyr58** downwards and to the left, thus being better buried in site X. This wasn't unexpected, since the sidechain of **Tyr58** was known to be mobile from previously studied structures. As a result, the crystal data (green) was closer to the docked fragment (blue) than to the docked final molecule (pink). In the case of **47**, there was perfect agreement between prediction and crystal data, validating once more the *in silico* work of K. Lal.

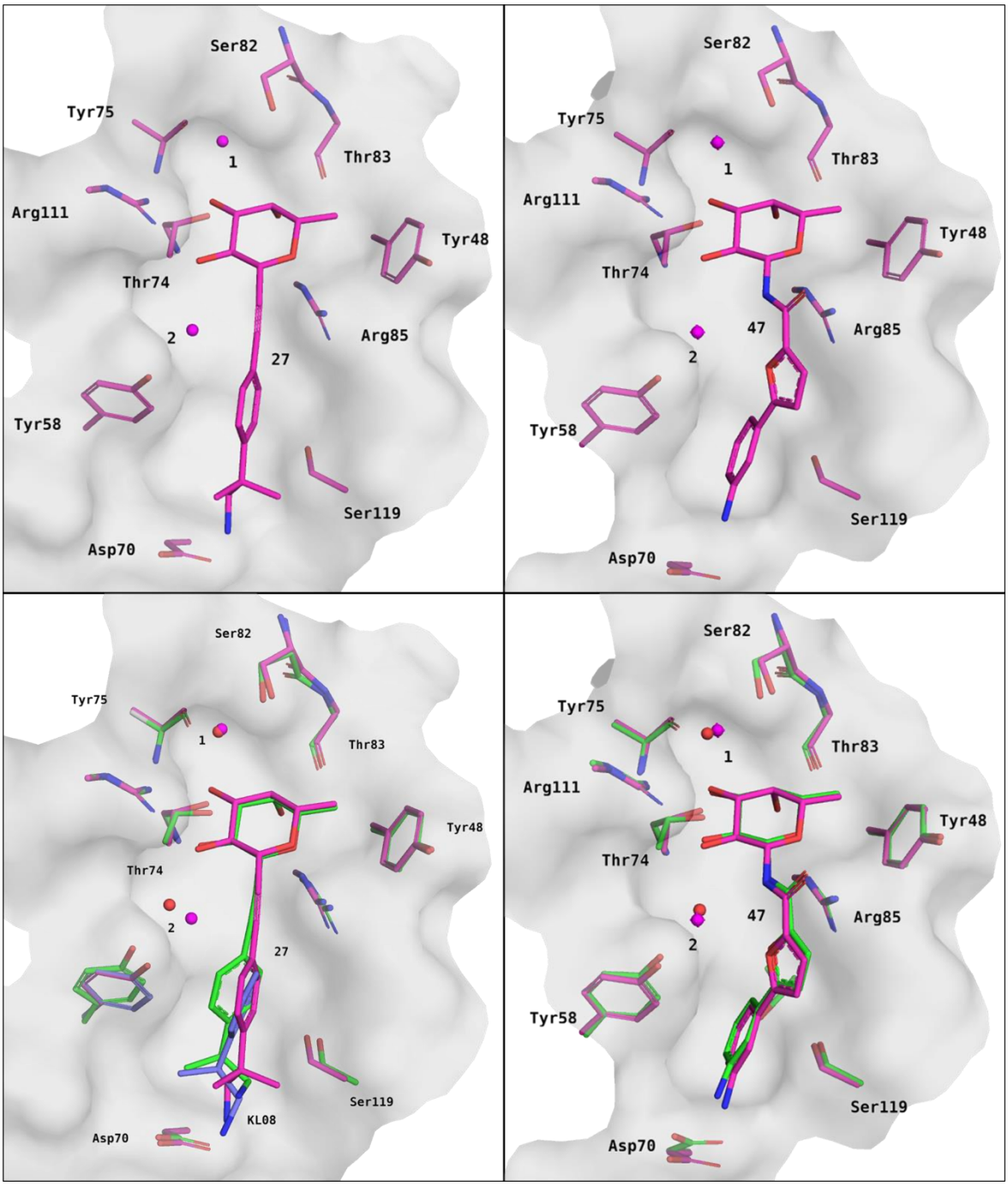


Figure 6.12. Top: Docking poses of ligands **27** (left) and **47** (right) on BC2L-C-N_{ter} (pink). **Bottom:** Superimposition with the new crystal structures (green) and docked fragment **KL08** (blue). Water molecules are depicted as red or pink spheres, protein surface in transparent gray.

In both structures, the fucose moiety conserved its known position, engaging in the previously described H-bonding with residues and crystallographic waters (**Figure 6.13**, left). In the structure featuring **27**, the alkyne linker was 4.1 Å long and didn't significantly displace the nearby crystallographic water **2**. The fragment moiety engaged in the predicted π/π T-shaped interaction with **Tyr58** at a distance of 3.9 Å, albeit the angle was *ca.* 60°, rather than 90°. Lastly, the salt bridge predicted by the docking pose of the fragment **KL08** (as seen in **Figure 5.2**) wasn't observed. This contact, involving the terminal amino group and **Asp70**, was

predicted for a fragment located 4.9 Å away from the anomeric position. Since the alkyne-bound **27** shrunk that distance by at least 0.8 Å, the amino group was too far from **Asp70** to be able to interact (4.4 Å). Nevertheless, a water-mediated contact between these groups was observed.

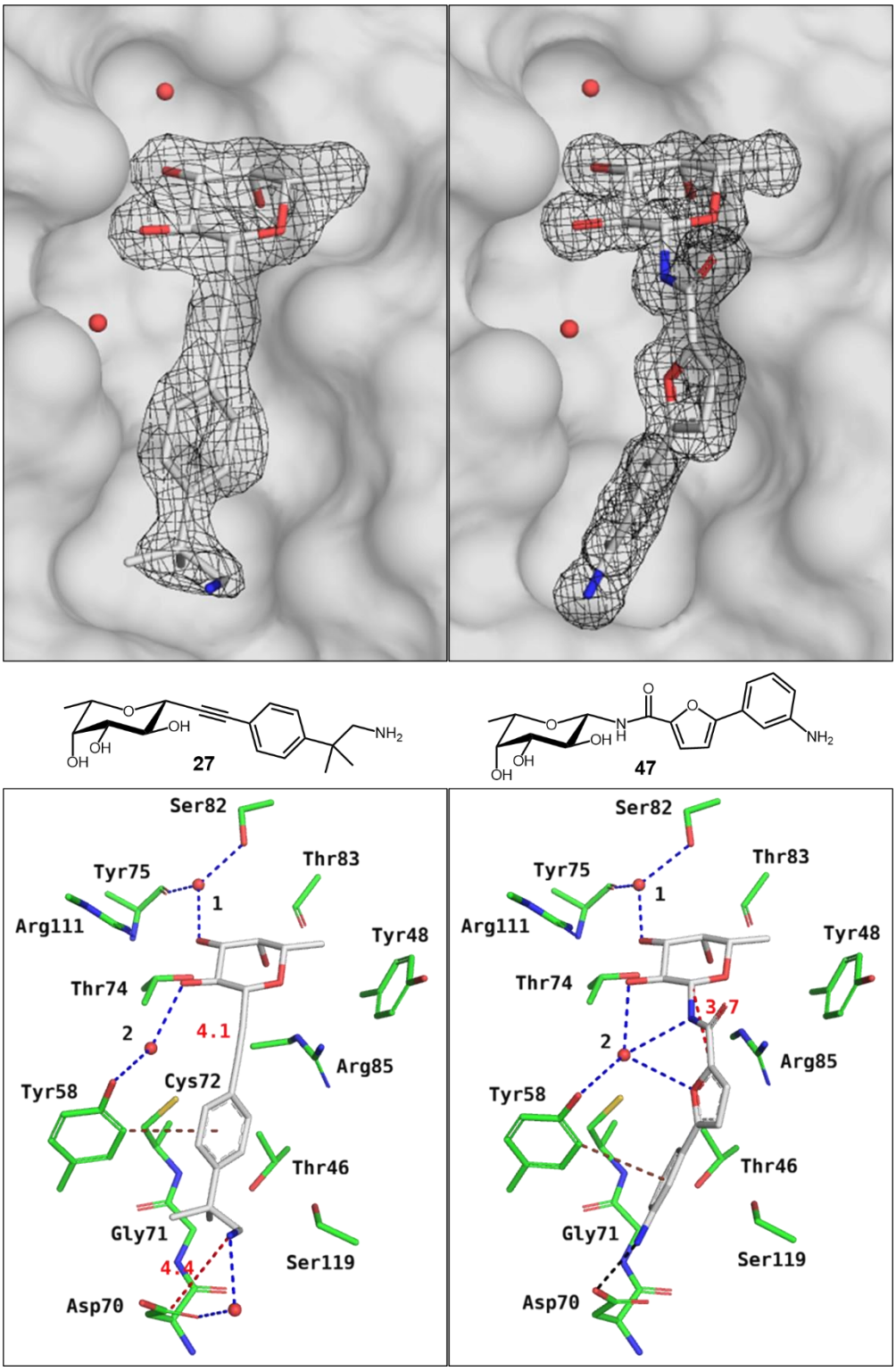


Figure 6.13. Top: Electronic density for synthetic ligands **27** (1.79 Å) and **47** (1.32 Å) in complex with BC2L-C-N_{ter}. New protein/ligand interactions are depicted in black or brown (hydrophobic), water contacts are depicted in blue. Distances (Å) from anomeric carbon to fragment or from amino group to **Asp70**, depicted in red.

Other characteristics of this interaction included the shape complementarity of hydrophobic patches: both methyl groups of the fragment were in close proximity with the otherwise exposed hydrophobic surface generated by the main chain of **Gly71** and side chain of **Tyr58**. This hydrophobic shape complementarity carried on to match residues **Ser119**, **Thr46** and **Cys72**.

On the other hand, the structure containing **47** featured an amide bond positioning the fragment moiety 3.7 Å away from the anomeric position (**Figure 6.14**, right). Designed to either replace or interact with crystallographic water **2**, the amide bond interacted through its nitrogen atom, while the carbonyl pointed towards the solvent. This interaction not only validated the purpose of using the amide bond, but also the *in silico* predictive docking. All predicted interactions were observed, including a π/π T-shaped interaction with **Tyr58** (3.6 Å) and H-bonds between water 2 and the furan moiety, as well as between residue **Asp70** and the aniline moiety (2.8 Å). In contrast with the other crystal structure, this last interaction is allowed by the length of the final molecule. It occurs as an H-bond instead of a salt bridge as the aniline moiety isn't charged at this pH. In terms of shape complementarity, this ligand is more solvent-exposed than the former, except for its aniline moiety, which matches the aforementioned hydrophobic patch composed of **Gly71**, **Tyr58**, **Thr46**, etc.

It is worth to note that across previous BC2L-C-N_{ter} structures, at least two water molecules consistently resided in the space now filled by the fragment moieties. This displacement of ordered water can translate into thermodynamic entropic gains, especially since the water molecules in question didn't establish consistent and conserved interactions with their surroundings, minimizing the potential enthalpic loss.

Altogether, the data presented confirms the compatibility of β-oriented substituents and the known fucoside binding mode. The alkyne and amide linkers are shown to be appropriate for this design. The structures also validate the predicted binding poses for ligand or fragment structures, the length of the linker being a limit for **27**. Additionally, we can rationalize the affinity gain observed for **27** as the result of three factors: (1) the T-shaped π/π interaction, (2) the shape complementarity between hydrophobic surfaces and (3) the thermodynamically advantageous entropic factor. Finally, it motivates the need to re-synthesize ligand **47** to attempt ITC measures and improve its poor solubility in a future 2nd generation design.

6.6. Outlook

We have explored a set of tools to evaluate BC2L-C-N_{ter} antagonists. Among the techniques used, STD-NMR and DSC showed promise as methods to quickly screen panels of new structures. This would allow validation of binding and early prioritization of ligands, all while remaining economical in terms of material expenditure. Nevertheless, these and other techniques were limited by the low affinity of BC2L-C-N_{ter} for monosaccharides, including the newly synthesized molecules. This was the reason for not pursuing FP, which may hold promise if a high-affinity fluorescent probe is developed for this lectin.

Similarly, SPR competition experiments were put aside due to the low affinity observed but were able to provide IC_{50} values that may allow ranking of fragments. Continuing with SPR, a first-ever BC2L-C-N_{ter} SPR chip was created with exceptionally long shelf life and that measured micromolar affinities accurately. Once more, millimolar affinities were difficult to assess confidently. Nevertheless, retrospective study indicated that SPR experiments allow predictions of ligand behaviour to some extent, which will improve as affinities of new structures do. The experimental protocol developed allows affinity assessment for low-yielding fragments and saves material for the final evaluation.

Indeed, ITC was material-costly but allowed unambiguous measure of ligand affinities in solution by the means of low c-value experiments. It led to the identification of two successful antagonists with improved affinity compared to the monosaccharide. The current leading structure **27** presented a 9-fold affinity gain and validated our ligand design strategy, as well as the use of alkyne linkers in antagonists. It has to be mentioned that the affinity of **27** is expected to improve when the regioisomeric mixture is eventually replaced by the pure *para*-compound. Interestingly, alkyne-bound ligands **16** and **27** presented curious thermodynamic profiles, to be further studied on future high-affinity alkyne antagonists. Finally, the first crystal structures of BC2L-C-N_{ter} complexes with synthetic ligands were solved, featuring antagonists **27** and **47**. This validated the computational and experimental work performed to date, as well as the use of amide linkers in antagonists.

Taken together, this evaluation campaign has laid the groundwork and experimental protocols for affinity and structural evaluation of antagonists. This will allow dynamic feedback for the synthetic efforts, which will be re-directed in function of new findings. For example, alkyne and amide linkers seem to perform better than triazoles, thus should be

prioritized. Similarly, molecule **27** is the current leading antagonist for BC2L-C-Nter, meaning derivatives should be produced in order to attempt SAR.

7. CONCLUSIONS AND PERSPECTIVES

In order to counter the growing threat of MDR pathogens, modern therapy is shifting its focus from antibiotics to alternative and complementary therapies. The use of anti-adhesion therapy for targeting virulence vectors has proven its worth, resulting in the development of glycomimetics targeting pathogenic lectins has gained traction in the last decades. In turn, the advent of glycomimetics has propelled efforts such as the PhD4GlycoDrug European Joint Doctorate.

Within the scope of the PhD4GlycoDrug consortium, the task had been established to target and antagonize the MDR pathogen *B. cenocepacia* through its family of BC2L-lectins. The work presented in this thesis consisted in the *Design, Synthesis and Evaluation of antagonists towards BC2L-C* and focused on the N-terminal lectin domain of this superlectin: BC2L-C-N_{ter}. As such, the project had three objectives:

- The production and characterization of a new construct for BC2L-C-N_{ter}
- The rational design and synthesis of fucoside antagonists for this target
- The evaluation of the produced glycomimetics against their target

All three of these objectives have been completed sequentially:

The first part of the thesis led to the production of a new, stable protein construct of BC2L-C-N_{ter} : rBC2L-CN2, which provided critical information on the still relatively uncharted N-terminal of the superlectin. Thanks to the new protein construct, any future work with BC2L-C-N_{ter} is facilitated: purification, characterization and crystallization are now facilitated. The new construct performed well in a wide range of techniques including SPR, ITC, STD-NMR, DSC, and crystallization. The affinities of BC2L-C-N_{ter} for human fucoside epitopes of the H-type family were fully assessed from the mono- to the hexasaccharide and rationalized both by affinity and structural evaluation. Three crystal structures featuring BC2L-C-N_{ter} complexed with oligosaccharide ligands Globo H, H-type 1 and Lewis y were obtained for the first time, proving the merit of the new construct. With the structural knowledge gathered and the use of ITC, the high affinity measured for these ligands was attributed to the third sugar unit and the interactions it establishes in the vicinity of the CRD.

The successes attained on the first part were translated to the second objective: the design and synthesis of fucomimetics targeting BC2L-C-N_{ter}. In collaboration with the *in silico* work of K. Lal, structures were designed to become lectin ligands. These structures featured both a fucose and a fragment moiety, the latter being obtained by computational screening of fragments against a CRD-vicinal site. Onwards, and with the contribution of mentored students, a synthetic framework was drafted and validated, which allowed the modular synthesis of β-C- and β-N-fucosides, linked by alkyne, alkene, amide, or triazole functions. The limiting factor for the synthetic project was the structural variability of the fragments considered. This limit found its solution with *de novo* synthesis, which increases the synthetic work but also allows access to a broader scope of structures for future synthesis. On this note, synthetic avenues opening the way to the design of a 2nd generation of ligands were briefly explored and should be pursued after the confirmation that the multivalent approach to antagonize this target is viable. Finally, thanks to this straightforward and broadly applicable synthetic framework, a panel of final molecules was obtained to be evaluated against their target: BC2L-C-N_{ter}. We now have the ability to generate libraries of glycomimetics with potential as lectin antagonists. Indeed, the modular synthesis of C- and N-fucoside glycomimetics will benefit this and other similar projects in the long term.

With the synthetic project established and running, the third objective was confronted: to establish a reliable system for the evaluation of BC2L-C-N_{ter} antagonists. Among other techniques, SPR and ITC protocols were developed to evaluate the affinity of the antagonists. A difficult limit to this evaluation was met in the low affinity of BC2L-C-N_{ter} for the synthetic structures: adapted experiments were designed and the results obtained were considered critically. Nevertheless, the system developed allows preliminary assessment of affinity by SPR, then reliable measure of the binding interaction by ITC. Moreover, techniques such as STD-NMR and DSC have shown their potential for becoming early screening and validation tools. The established evaluation system is expected to continue for future generations of BC2L-C-N_{ter} antagonists and rely broader and more reliable data as their affinity increases. Indeed, measuring higher affinities by ITC will provide thermodynamic insight into the binding event, enabling the rationalization of affinity gains in terms of entropy and enthalpy and allowing enhanced ligand design. Finally, crystallographic work has led to solving two crystal structures of BC2L-C-N_{ter} complexes with synthetic ligands, which provided critical structural data and served as a validation for the entirety of the work performed in this project, including

the *in silico* work. This is an exciting fact since it showcases the potential and accuracy of computational methods for antagonizing therapeutic target. The evaluation of the 1st generation of antagonists has provided important feedback: promising structures feature alkyne and amide bond linkages, with fucomimetic **27** (see **Figure 7.1**) becoming the leading antagonist of the project with a 9-fold affinity increase over its parent structure.

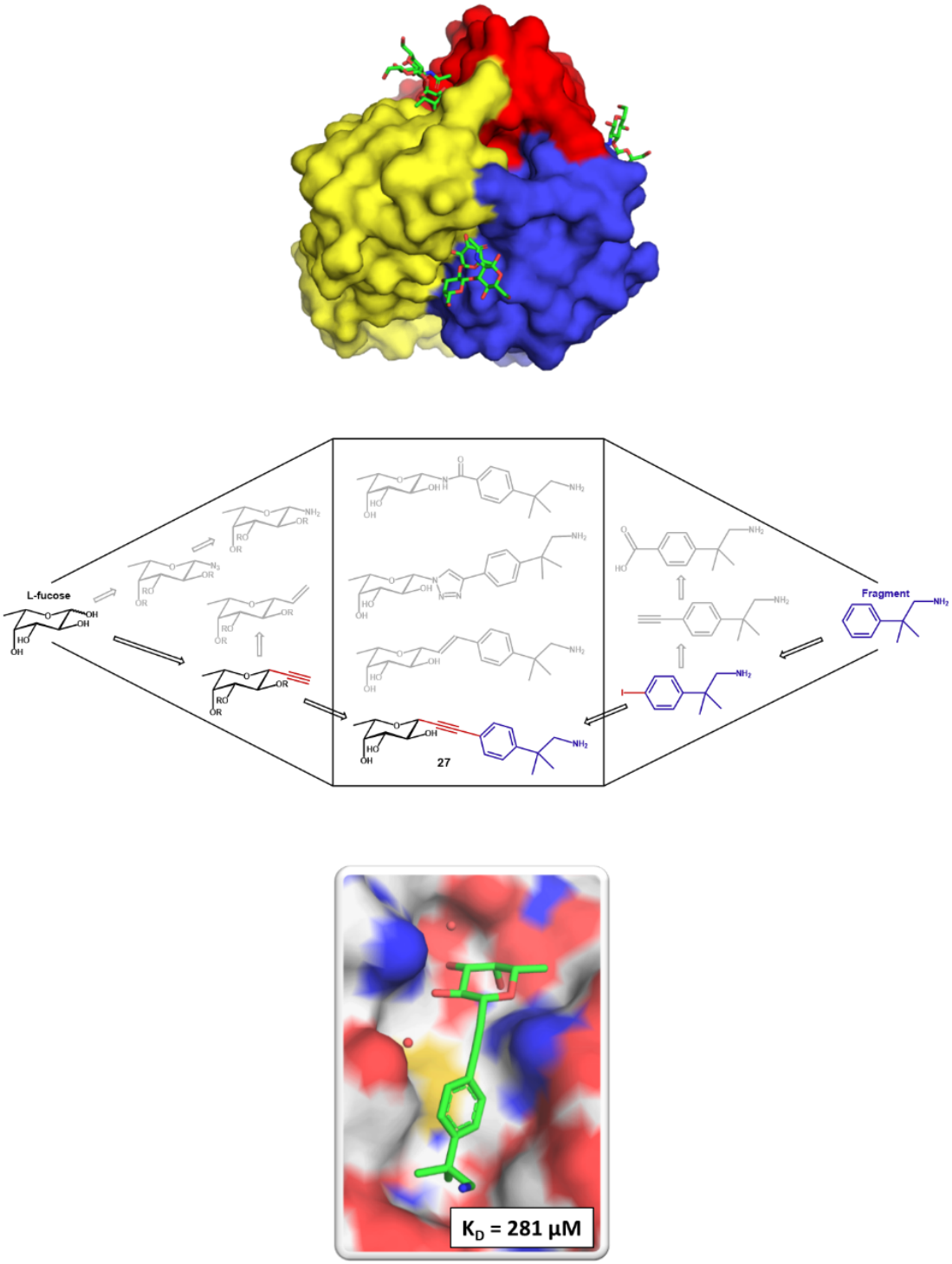


Figure 7.1. Design, synthesis and evaluation of BC2L-C antagonists.

Thus, the necessary tools have been created for long term targeting of BC2L-C-N_{ter}: a reliable protein construct, a modular synthetic framework and comprehensive evaluation platform. Overall, this successful campaign has opened the way to effectively probing and drugging BC2L-C-N_{ter}, as well as other yet undiscovered lectins with an equivalent binding mode. Onwards, it will be interesting to test the leading antagonist and subsequent structures in functional assays such as cell-adhesion, hemagglutination and biofilm formation. Indeed, learning more about BC2L-C, its interactions with human epitopes, and how to antagonize them is certainly a relevant step in the larger endeavour against *B. cenocepacia* and other related pathogens presenting similar virulence factors.

In broader terms, this project has been a good example of how to study and target a therapeutic target. The workflow employed, tools used, and experience gained will be useful for future projects undertaken in the world of medicinal chemistry. Although this can be considered a small contribution to the endeavour of science against pathology, it allowed to validate the worthiness of the strategy taken and reveal some of the classic pitfalls of such projects. Interdisciplinary work was certainly essential to examine this project from all its angles and collaboration and communication were invaluable tools for broadening the scientific scope of the work. Altogether, the *Design, Synthesis and Evaluation of antagonists towards BC2L-C* has proven to be an excellent training for a scientist looking forward to employ chemistry and biotechnology to take part in humanity's efforts against pathologies.

8. APPENDIX

8.1. Scientific Communication: secondment at Glycopedia

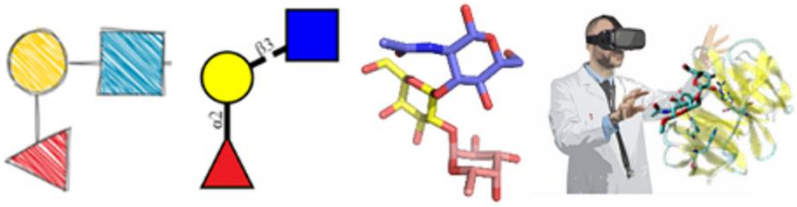
As stated in Section 2, PhD4GlycoDrug established a goal of disseminating glycoscience freely, in coordination with Glycopedia.[163] Consequently, a secondment was programmed at Glycopedia during the spring of 2020. During this time, and in collaboration with Kanhaya Lal and Dr. Serge Perez, a chapter was drafted to be added to Glycopedia. It went on to become an open-access review article in the Beilstein Journal of Organic Chemistry.[164] 'Computational tools for drawing, building and displaying carbohydrates: a visual guide' is presented in its entirety in the following section.

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

OPEN ACCESS PLATINUM

Computational tools for drawing, building and displaying carbohydrates: a visual guide

Kanhaya Lal, Rafael Bermeo and Serge Perez



Beilstein J. Org. Chem. **2020**, *16*, 2448–2468. https://doi.org/10.3762/bjoc.16.199

CC BY 4.0

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Computational tools for drawing, building and displaying carbohydrates: a visual guide

Kanhaya Lal‡[1,2], Rafael Bermeo‡[1,2] and Serge Perez*[1]

Review

Open Access

Address:
[1]Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France and [2]Dipartimento di Chimica, Università Degli Studi di Milano, via Golgi 19, I-20133, Italy

Email:
Serge Perez* - spsergeperez@gmail.com

* Corresponding author ‡ Equal contributors

Keywords:
bioinformatics; carbohydrate; glycan; glycobiology; nomenclature; oligosaccharide; polysaccharide; representation; structure

Beilstein J. Org. Chem. **2020,** *16,* 2448–2468.
https://doi.org/10.3762/bjoc.16.199

Received: 18 June 2020
Accepted: 17 September 2020
Published: 02 October 2020

This article is part of the thematic issue "GlycoBioinformatics".

Guest Editor: F. Lisacek

© 2020 Lal et al.; licensee Beilstein-Institut.
License and terms: see end of document.

Abstract

Drawing and visualisation of molecular structures are some of the most common tasks carried out in structural glycobiology, typically using various software. In this perspective article, we outline developments in the computational tools for the sketching, visualisation and modelling of glycans. The article also provides details on the standard representation of glycans, and glycoconjugates, which helps the communication of structure details within the scientific community. We highlight the comparative analysis of the available tools which could help researchers to perform various tasks related to structure representation and model building of glycans. These tools can be useful for glycobiologists or any researcher looking for a ready to use, simple program for the sketching or building of glycans.

Introduction

Glycoscience is a rapidly surfacing and evolving scientific discipline. One of its current challenges is to keep up and adapt to the increasing levels of data available in the present scientific environment. Indeed, the rise of accessible experiment data has changed the landscape of how research is performed. The accessibility of this information, coupled with the emergence of new platforms and technologies, has benefitted glycoscience to the point of enabling the detection and high-resolution determination and representation of complex glycans [1]. Increasing numbers of carbohydrate sequences have accumulated throughout extensive work in areas of chemical and biochemical fragmentations followed by analysis using mass spectroscopy, nuclear magnetic resonance, crystallography and computational modelling. There have been some initiatives by independent research groups worldwide, that pushed the development of visual tools to improve some aspects of glycan identification, quantification and visualisation, some of which will be further developed throughout this article.

Biological molecules express their function throughout their three-dimensional structures. For this reason, structural biology places great emphasis on the three-dimensional structure as a central element in the characterisation of biological function. An adequate understanding of biomolecular mechanisms inherently requires our ability to model and visualise them. Visualisation of molecular structures is thus one of the most common tasks performed by structural biologists. As an essential part of the research process, data visualisation allows not only to communicate experimental results but also is a crucial step in the integration of multiple data derived resources, such as thermodynamics and kinetic analysis, glycan arrays, mutagenesis, etc. Data visualisation remains a challenge in glycoscience for both the developers and the end-users even for the simple task of describing molecular structures. Progress in this area allows to translate a static visualisation of single molecules into dynamic views of complex interacting large macromolecular assemblies, which increases our understanding of biological processes.

Representing the structures of carbohydrates has historically been considered to be a complicated task. Starting from the linear form of the Fischer projection, which is certainly not a realistic representation of a carbohydrate structure, there has been a continuous development and evolution of the description of monosaccharides [2]. Glycans are puzzles to many chemists, and biologists as well as bioinformaticians. This complexity occurs at different levels (which makes it incremental). Amongst the most recognisable "sugars", glucose is merely one of 60+ monosaccharides, all of which are, in truth, pairs of mirror-image enantiomers (L and D).

Moreover, monosaccharides occur as two forms: 5-atom ring (furanose) and 6-atom ring (pyranose). With the occurrence of a statistically rarer "open form," we obtain at least 6 "correct" representations of glucose. And yet, monosaccharides are only the chemical units and the individual building blocks of much more complex molecules; the carbohydrates, also referred to as glycans. The glycan family can be grouped in the following categories: (i) oligosaccharides (comprising two to ten monosaccharides linked together either linearly or branched); (ii) polysaccharides (for glycan chains composed of more than ten monosaccharides); (iii) glycoconjugates (where the glycan chains are covalently linked to proteins (glycoproteins), lipids (glycolipids). The complexity of glycans is a consequence of their branched structure and the range of building blocks available. Other levels of complexity include the nature of the glycosidic linkage (anomeric configuration, position and angles), the number of repeating units (polysaccharides) as well as the substitutions of the monosaccharides. Regardless of the different nomenclatures available to describe each monosaccharide, representing and encoding a glycan structure into a file is required for communication among scientists as well as for data processing.

As a consequence, glycobiologists have proposed different graphical representations, with symbols or chemical structures replacing monosaccharides. The description of carbohydrate structures using standard symbolic nomenclature enables easy understanding and communication within the scientific community. Research groups working on carbohydrates have developed schematic depictions with symbols [3] and expansions with greyscale colouring as the so-called Oxford nomenclature (UOXF) [4,5], and even fully coloured schemes later on. Among these, some of the proposed representation forms have been accepted and implemented by several groups and initiatives, namely the Consortium for Functional Glycomics (CFG) [6]. Whereas the initial versions of such representation were limited to mammalian glycans, an extension of the graphical representation of glycans, called SNFG Symbol Nomenclature for Glycans (SNFG) [7,8] resulted from a joint international agreement. The newly proposed nomenclature covers 67 monosaccharides aptly represented in eleven shapes and ten colours. There is the hope that it will cope better with the rapidly growing information on the structure and functions of glycans and polysaccharides from microbes, plants and algae. The rendering of glycan drawing and symbol representations motivated the development of several computer applications using a standardised notation. The earliest glycan editors allowed manual drawing similar to ChemDraw or used input files with glycan sequence KCF (KEGG Chemical Function) [9] in text format for similarity search against other structures deposited in the databases. Later developments supported the construction and representation of glycan structures in symbolic form by computational tools like GlycanBuilder [10]. Since then, several advancements have been made to allow the user to both draw glycans manually or by importing and exporting the structure files in different text formats [11].

Along the same line, the development of various other applications allowed the users to sketch 2D-glycan structures by dragging and dropping monosaccharides to canvas to generate 3D structures for further usages. These depictions comply with protein data bank (PDB) [12] format, or in the form of images [13,14]. Besides, these tools for representing glycans in 2D and 3D shape [15] allowed the integration of glycans into protein structures or complexes. The tools developed in the last few years have automated the sketching of glycans and glycopeptides, allowing rapid display of structures using IUPAC format [16] as input. This article explores and illustrates the concepts of "sketching", "building" and "viewing" glycans (Figure 1). It provides a descriptive analysis of the tools available for such

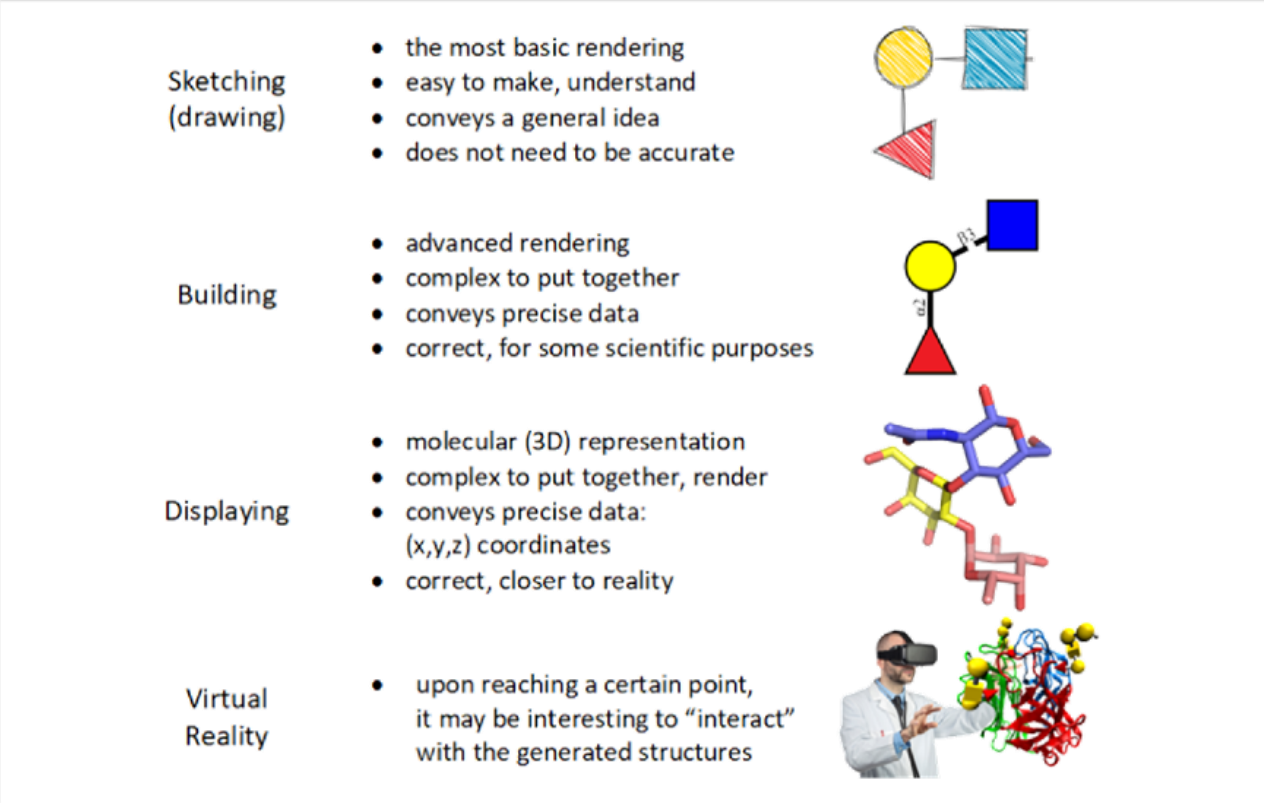


Figure 1: Levels of representation of glycans: from sketching to virtual reality.

activities, which can be useful for researchers looking for a ready-to-use simple program for sketching, building and 3D structure analysis of glycans and glycoconjugates. The scope of this work is relevant to N- and O-linked glycans, glycolipids, proteoglycans and glycosaminoglycans, lipopolysaccharides, plant, algal and bacterial polysaccharides.

Review

Methods

To facilitate glycoscience research, we have identified the tools and databases that are freely available on the internet and are regularly updated and improved [1]. The variety and complexity of glycan structures make their interpretation challenging. Consequently, in the past few years, several sketching, building and visualisation tools have been developed to depict better and understand the complex glycan structures. In this study, the freely available tools have been visited (April 2020) and analysed to highlight their core features but also explore their unique advancements to facilitate glycan research. Each of the computational tools was inspected for general features related to sketching, representing and model building, all of which could be further used as input for translation into other formats, search from glycan databases or complex calculations such as molecular simulations. Several tools feature an interactive interface which allows for manual editing of the structures. Examples of such tools are DrawRINGS [17], KegDraw [18], Glycano (available at http://glycano.cs.uct.ac.za/), GlycoEditor [19], GlycanBuilder [20], etc. These tools (except KegDraw) are provided with the list of CFG symbols to freely build glycan structures using the mouse on the canvas. In addition to manual sketching, some of these tools also can import text formats including IUPAC-condensed, GlycoCT and KEGG Chemical Function (KCF) format to display the glycan structures. Some applications also facilitate glycan search in various databases. Another category of tools included in this study involved glycan viewers which can only depict structures using the IUPAC three letter code or IUPAC-condensed nomenclature as input. These tools convert the input into a 2D image or 3D representation using SNFG symbols or 3D-SNFG illustration. Additionally, 3D representation of structures is provided by tools such as Visual Molecular Dynamics (VMD) [21] , and LiteMol [22], which allow for quick analysis of structural features in 3D space. All the tools mentioned were evaluated against a set of pre-selected criteria relating to ease of use, scientific precision and content, among others.

Table S1 (Supporting Information File 1) schematically summarises how these criteria are fulfilled. The analysis of the tools for input and output formats also provided information about their versatility to convert results into the standard or

desired format. The tools have been attributed to categories such as "sketcher", "builder" and "viewer", with eventual overlaps. A brief analysis of each application ordered by category is given in the next section.

Sketching with the free hand

As a preview of the following parts of this study, we performed an initial test of the tools available for the representation of a simple disaccharide: lactose (β-D-Gal*p*-(1→4)-D-Glc*p*).

Figure 2 shows how different web-available platforms rendered it. On the one hand, thanks to the unified nomenclature, there is no ambiguity regarding the nature of the carbohydrate represented. On the other hand, small differences between sketches appear. Such variations will multiply with the increasing complexity of the carbohydrates. It is, therefore, essential to choose which tools to use before starting an hour-long "drawing-spree". The variations of the colour code used to represent the monosaccharides show striking differences across platforms even though the appropriate colours to be used are strictly defined (https://www.ncbi.nlm.nih.gov/glycans/snfg.html#tab2). The colour discrepancy observed here means that some of the tools do not conform to SNFG standards. For some purposes, this conformity might not be a strict necessity. Another pronounced disparity concerns the representations of the glycosidic linkage. Across sketches the length/width of the linkage varies, which will result in either compact or extended images, to be taken into account when considering the size available for the intended figures. Finally, the sketches provide further information about the linkage type: anomericity and position. These details can be either useful or superfluous depending on what is the intended use for the finished design. The main characteristic of a helpful sketching tool should be its adaptability. By allowing to modify colours, sizes, lengths/widths and turn some features on/off, a "sketcher" would allow maximum flexibility to depict carbohydrates in any desired or necessary form, size, orientation. However, this adaptability should become available without hampering the sketching effort. The perfect sketching tool would, therefore, combine flexibility and high usability.

Building with scientific accuracy

The necessity for precision is what, at some point, turns carbohydrate sketching into building. What defines this turning point (besides a certain level of accuracy) is the intended purpose for the produced figures/images. Scientific communication, comparison between similar yet different structures, or merely showcasing the complexity of carbohydrates: all three cases cannot rely on a sketching tool to convey their message. Consequently, a new set of considerations appears. The requirement for accurate depiction comes from the complexity mentioned

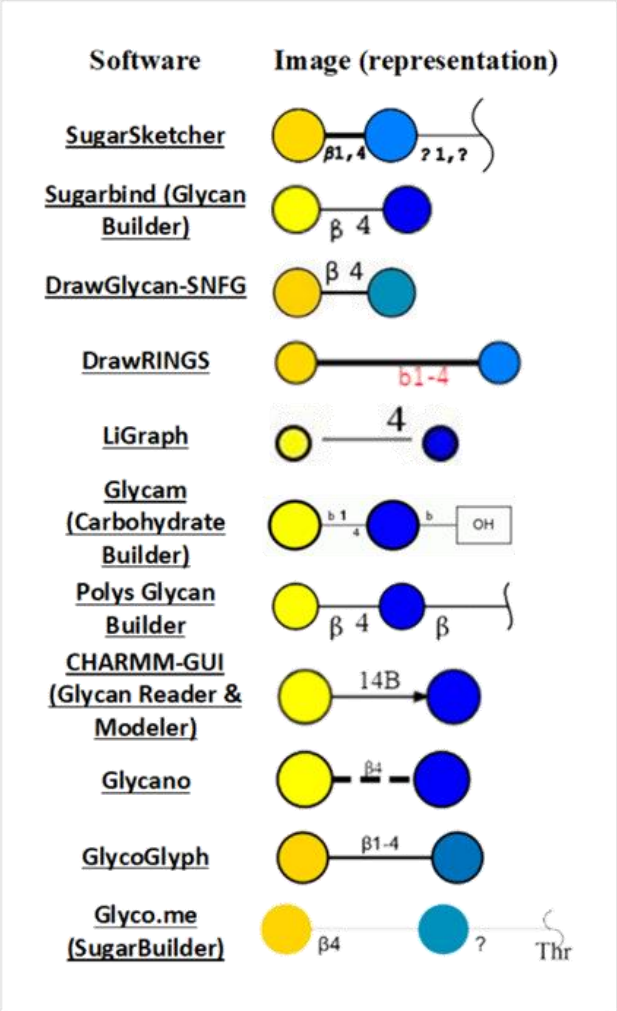


Figure 2: Depiction of lactose by various glycan sketching tools.

above of carbohydrates: anomeric configuration, substitution, glycosidic bond position, and repeating units (as well as tethering to larger macromolecules, and more). For the sake of accuracy, only the right combination of characteristics should be depicted, leaving no ambiguity: every relevant piece of data should be detailed. The glycosidic linkage is a perfect example to illustrate the necessity for accuracy in building, as opposed to sketching. While a simple line is enough to link two monosaccharides, it is necessary to define the linkage as alpha or beta (or unknown) and to state the positions of the glycosyl acceptor and even donor. Cellulose and amylose are two glucose-based polysaccharides that differ only in the nature of their glycosidic bond, and yet they have entirely different shapes and so, biological roles. For the sake of completion, the full description of a monosaccharide should obey the following rules: *<anomeric prefix><prefix for absolute configuration><the monosaccharide code><prefix for ring configuration>[<O-ester and O-ether substitutions and positions>]*. It is thus necessary to include such information when depicting carbohydrates, but

such features are simply absent in most of the existing glycan sketching tools.

Another feature that may become essential when the carbohydrate at hand is a polysaccharide is the possibility of building repeating units. Without this option, it would be simply impossible to build the required depiction. It emerges that an efficient carbohydrate builder must offer a wide array of options to characterise and personalise each monosaccharide. This would, in turn, entail a multitude of buttons, switches, etc.; which would result in a very complex interface. Consequently, unless the interface is rather straightforward and the building dynamic is well-designed, the software would be too difficult to use effectively. The ideal carbohydrate builder pick would also allow liberty for the user in terms of levels of precision since it has to fit every level of complexity above sketching. Lastly, once the building process is complete, a good builder must not only render all the provided data in the form of a precise figure but also allow the transfer of the data to other platforms (for example, by exporting the generated code).

Force fields for carbohydrates, 3D model building and beyond

Carbohydrates present various challenges to the development of force fields [23]. The tertiary structures of monosaccharides usually have a high number of chiral centres which increases the structural diversity and complexity. The structural diversity changes the electrostatic landscape of molecules; thus, it provides challenges in the development of force fields for accurate modelling of such variations in charge distributions. The monosaccharides can further form a large number of oligosaccharides which can enormously increase the conformational space, due to a high number of rotatable bonds. Nonetheless, recent developments in carbohydrate force fields enable to model and reproduce the energies associated with minute geometrical changes. The currently available force fields which are parameterised for carbohydrates are also capable of carrying out simulations of the oligosaccharides containing additional groups like sulfates, phosphates etc. [24] Generally used force fields for the Molecular Mechanics (MD) simulation of carbohydrates are CHARMM [25], GLYCAM [26], and GROMOS [27]. The structural complexity increases the computational cost, which makes simulations of large systems more challenging. Therefore, coarse-grained models [28] for carbohydrates are generally used for molecular modelling of large systems.

In terms of 3D model building, the complex topologies of glycans require dedicated molecular building procedures to convert sequence information into reliable 3D models. These tools generally use 3D molecular templates of monosaccharides to reconstruct a 3D model. Energy minimisation methods can further refine the models. These models are essential for structure-based studies and complex calculations like Molecular Dynamics simulations. Therefore, the accurate model building requires the use of reliable databases to generate atomic coordinates and topology to provide an acceptable model. Some of the computational tools usually contain atom coordinates of generally used monosaccharides (as templates) and also use libraries of bond and angle parameters from various force fields dedicated for carbohydrates. The accurately predicted oligosaccharide conformations are good starting points for further investigations. Of particular interest are the evaluations of the dynamics of glycans and their interactions with proteins which is a most significant concern in glycoscience. The joint need to better perceive and manipulate the three-dimensional objects that make up molecular structures is leading to a rapid appropriation of techniques of Virtual Reality (VR) by the molecular biology community. Generic definitions describe VR as being immersion in an interactive virtual reactive world. The computer-generated graphics provide a realistic rendering of an immersive and dynamic environment that responds to the user's requests. One finds in these definitions the three pillars that define VR: Immersion, Interaction, Information. Although it is difficult to extract a single, simple definition of VR, the main idea is to put the user at the centre of a dynamic and reactive VR environment, artificially created and which will supplant the real world for the time of the experiment.

Input and output for sketching, building and displaying applications

The variety and complexity of carbohydrate structures hamper the use of a unique nomenclature. The choice of notation depends on whether the study is focused on chemistry or has a more biological approach. The IUPAC-IUBMB (International Union for Pure and Applied Chemistry and International Union for Biochemistry and Molecular Biology) terminologies, in their extended and condensed forms [16], govern the naming of the primary structure or sequence.

Further down the line, the complexity of the existing nomenclatures for carbohydrate-containing molecules remains a significant hurdle to their practical use and exchanges within and outside the glycoscience cenacle. The linearisation of the description of the structure is a way to cope with the description of the structural complexity. The proposed formats provide rules to extract the structure of the branches and create a unique sequence for the carbohydrate. The most commonly used formats are IUPAC [16], GlycoCT [29], KCF [9], and WURCS [30].

The sketching of carbohydrates using computational tools generally requires the textual input and output in at least one of

Input Output formats	
IUPAC condensed	Man(a1-3)[Man(a1-6)]Man(b1-4)GlcNAc(b1-4)b-GlcNAc
LINUCS	[]\[b-D-GlcpNAc\]{[(4+1)][b-D-GlcpNAc]{[(4+1)][b-D-Manp]{[(3+1)][a-D-Manp]{}[(6+1)][a-D-Manp]{}}}}
GlycoCT	RES 1b:b-dglc-HEX-1:5 2s:n-acetyl 3b:b-dglc-HEX-1:5 4s:n-acetyl 5b:b-dman-HEX-1:5 6b:a-dman-HEX-1:5 7b:a-dman-HEX-1:5 LIN 1:1d(2+1)2n 2:1o(4+1)3d 3:3d(2+1)4n 4:3o(4+1)5d 5:5o(3+1)6d 6:5o(6+1)7d
KCF	ENTRY G12157 Glycan NODE 6 1 Asn 18 0 2 GlcNAc 9 0 3 GlcNAc -1 0 4 Man -11 0 5 Man -19 3 6 Man -19 -3 EDGE 5 1 2:b1 1:4 2 3:b1 2:4 3 4:b1 3:4 4 5:a1 4:6 5 6:a1 4:3 ///
WURCS	WURCS=2.0/3,5,4/[a2122h-1b_1-5_2*NCC/3=O][a1122h-1b_1-5][a1122h-1a_1-5]/1-1-2-3-3/a4-b1_b4-c1_c3-d1_c6-e1

Figure 3: Examples of different glycan structure text formats for the same glycan. Data in these formats are generally used as input/output in glycan drawing and 3D structure building tools.

these formats (Figure 3). An alternate input method involves manual sketching of 2D glycan structures by dragging and dropping monosaccharide symbols on canvas (with or without grids) to connect them further. This method makes the sketching tools more friendly and interactive as it does not require large text code as input. Both input methods are compliant to the Symbol Nomenclature for Glycans (SNFG). Another symbolic representation that could clearly distinguish monosaccharides in monochrome colours is the Oxford notation [5]. In this method, dashed and solid lines represent the alpha and beta glycosidic linkages, respectively. There are few tools which have implemented this method while other tools use text to highlight this information in the structures. In addition to sketching tools, some applications, specific to the field of carbohydrates, provide the possibility to visualise and display 3D structures. These visualisation tools accept strings or files in text formats (GlycoCT, IUPAC-condensed, KCF) to display the structure via a graphical user interface. For instance, the DrawGlycan-SNFG [31] tool uses IUPAC-condensed nomenclature for input string and converts it into a 2D image represented in SNFG symbols. At the same time, the 3D-SNFG [15] can generate glycan structures by incorporating SNFG symbols in 3D space for further visualisation using the computational tools like visual molecular dynamics (VMD) [21] LiteMol [22] and Sweet Unity Mol [32].

Glycan sketchers

SugarSketcher. SugarSketcher [14] is a JavaScript interface module currently included in the tool collection of Glycomics@ExPASy (available at https://glycoproteome.expasy.org/sugarsketcher/) for online drawing of glycan structures. The interactive graphical interface (Figure 4, top) allows glycan drawing by glycobiologists and non-expert users. In particular, a "Quick Mode" helps users with limited knowledge of glycans to build up a structure quickly as compared to the normal mode, which offers options related to the structural features of complex carbohydrates (for example additional monosaccharides, isomers, ring types, etc.). The building of glycan structures uses mouse and proceeds via a selection of monosaccharides, substituents and linkages from the list of symbols. However, some wrong combinations of choices can block the interface, resulting in the need to re-start the process (SugarSketcher is on version beta 1.3). Alternatively, SugarSketcher also uses GlycoCT or a native template library as an input. A list of pre-built core N- and O-linked carbohydrate moieties, which are usually present in glycoproteins structures, can be used as a template for further modification. A shortlist of glycan epitopes is also included providing templates for drawing more complex molecules. The software uses the Symbol Nomenclature for Glycans (SNFG) notation for structure representation and exports the obtained sketch to text format (GlycoCT) or image (.svg) files. The software SugarSketcher is featured in the web portal GlyCosmos (https://glycosmos.org/glytoucans/graphic) [33]. Under the name "SugarDrawer", it provides an interface for generating carbohydrate structures to query the database included in GlyCosmos: GlyTouCan [34].

GlyCosmos is a web portal that integrates resources linking glycosciences with life sciences. Besides elements such as "SugarDrawer" and GlyTouCan (carbohydrate database), the platform GlyCosmos assembles data resources ranging from glycoscience standard ontologies to pathologies associated with glycans. GlyCosmos is recognized as the official portal of the Japanese Society for Carbohydrate Research and provides information about genes, proteins, lipids, pathways and diseases.

GlyTouCan (Figure 5) is a repository for glycans which is freely available for the registry of glycan structures. The repository can register structures ranging from monosaccharide compositions to fully defined structures of glycans. It assigns a unique accession number to any glycan to identify its structure and even allows to know its ID number in other databases. Al-

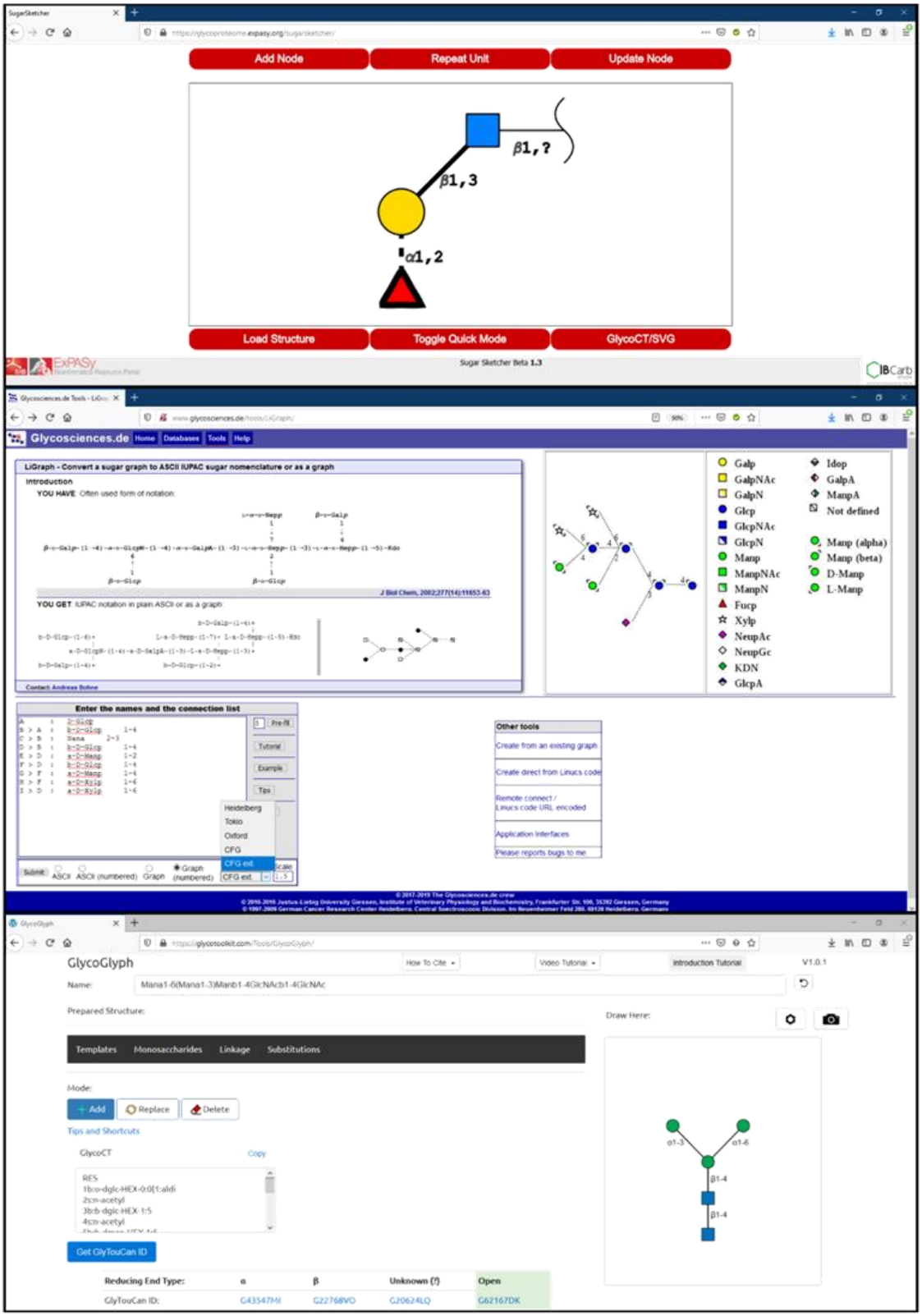


Figure 4: From top to bottom: SugarSketcher [36] interface with a glycan structure drawn using the "Quick Mode". LiGraph interface showing input and output options for glycan structure representation. GlycoGlyph [37] interface with a text input (modified IUPAC condensed) converted into its glycan image.

Figure 5: GlyTouCan [38] interface allows to search for glycans structures in the database. Data contained in GlyCosmos portal (https://glycosmos.org/) and in GlyTouCan repository home page (https://glytoucan.org/), including their logos, are licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

ternatively, users can search and retrieve information about the glycan structures and motifs that have been already registered into the repository. The structures can be searched simply by browsing through the list of already registered glycans or by specifying a particular sub-structure to retrieve structurally similar glycans (https://glytoucan.org/Structures/graphical). The software tool featured in the GlyTouCan website is called GlycanBuilder and is presented in a later section of our analysis.

Recapitulating, SugarSketcher can be an efficient tool for non-glycobiologists or glycobiologists to sketch glycans. However, it does not accept different input or output formats like IUPAC, WURCS (Web3 Unique Representation of Carbohydrate Structures), which would make the tool more versatile.

LiGraph. LiGraph [35] (http://www.glycosciences.de/tools/LiGraph/) is an online tool based on the concept of schematic drawings of oligosaccharides to display glycan structures. This tool also renders images of glycans in different notation using a text input. The input for the carbohydrate structure consists of a list of names and connections. The glycan structure is output in the specified notation: either ASCII IUPAC sugar nomenclature or a graph which can be rendered in different themes which include Heidelberg, Oxford, Tokyo, CFG and extended CFG (Figure 4, middle). The output images for the glycan structure and the legends can be saved and downloaded in .svg format. This tool is useful for glycan sketching using text templates, but its shortcomings include a limited number of monosaccharide symbols and restricted compatibility with other input file formats.

GlycoGlyph. GlycoGlyph [39] is a web-based application (available at https://glycotoolkit.com/Tools/GlycoGlyph/) built using JavaScript which allows users to draw structures using a graphical user interface or via text string in the CFG linear (also known as modified IUPAC condensed) nomenclature dynamically. The interface (Figure 4, bottom) is equipped with templates for N- and O-linked glycans and terminals. Also, it provides 80+ monosaccharide (SNFG) symbols and a selection for substituents. The selected template or text string (in CFG linear nomenclature) input directly gets converted into an image in canvas and also appears as text in GlycoCT format. The output can be saved as a .svg file or as GlycoCT text. The interface also provides additional options to add, replace or delete each monosaccharide, modify the sizes of symbols and text

fonts, and turn off the linkage annotations or change their orientation; all of which increases the usability of the software. The input structure can be further used to search the GlyTouCan [34] database to explore the literature details related to the input structure.

GlycoGlyph is an efficient tool for sketching or building glycans with a highly usable interface that can significantly help researchers to improve the uniformity in glycan formats in literature/manuscripts. It can also be a tool of choice for text mining for the query structure.

GlycanBuilder2. GlycanBuilder2 [40] is a Java-based glycan drawing tool which runs locally as an application on different platforms including Windows, macOS and Linux. It is freely available for downloading at http://www.rings.t.soka.ac.jp/downloads.html. GlycanBuilder2 is a newer version of GlycanBuilder [20] with additional features. This version is capable of supporting various ambiguous glycans consisting of monosaccharides from plants and bacteria. The tool uses the SNFG notation to display glycan structures. Moreover, this updated version can convert a drawn structure into WURCS sequences for further use as a query for glycan search or registration in databases like GlyTouCan. GlycanBuilder2 provides an excellent interface (Figure 6, top) for glycan drawing. Glycan structures can be drawn manually using the mouse or by importing text input files. The interface provides a list of templates: N- O-glycans, glycosphingolipids, glycosaminoglycans(GAGs). Rows of CFG notations for monosaccharides assist with glycan structure drawing on canvas. The application also supports the glycan symbol notations for the University of Oxford (UOXF) format. The input complies with various linear sequence and text formats. They include GlycoCT, GLYcan structural Data Exchange using Connection Tables (GLYDE-II), Bacterial Carbohydrate Structures DataBase (BCSDB) [41], carbohydrate sequence markup language (CabosML) [42], CarbBank [43], LinearCode [44], LINUCS, IUPAC-condensed and GlycosuiteDB [45]. The output yields structures in the following formats: GlycoCT, LinearCode, GLYDE-II and LINUCS. Thus, GlycanBuilder2 is a versatile tool which can be used for glycan sketching or building and also as a glycan sequence converter from one format to another.

Original GlycanBuilder. GlycanBuilder [10,20] was originally part of the GlycoWorkbench platform [49]. This interface is integrated in most tools of the Glycomics@ExPASy collection that require a drawing interface to query data. GlycanBuilder is written in Java Programming language and can be used as standalone or as an applet for embedding in web pages for glycan search. For example, GlycanBuilder is integrated in SugarbindDB [50] to draw glycan structures and search the database (https://sugarbind.expasy.org/builder), and in GlycoDigest or GlyS3 [10,20] to define the input of these tools.

Technically, the tool provides an interactive interface which allows an automated glycan rendering using a library of individual monosaccharides or pre-built template structures (Figure 6, middle). GlycanBuilder provides access to 41 templates. They include N- and O-linked glycans, GAGs (glycosaminoglycans), glycosphingolipids and milk oligosaccharides. It also contains 68 entries from MonosaccharideDB (http://www.monosaccharidedb.org/) including monosaccharides, modifications (e.g. deoxy) and substituents. The tool provides options to modify a monosaccharide by adding substituents and alterations. Free movement of the monosaccharides is allowed through movement and orientation buttons. GlycanBuilder offers multiple options for glycan notation which include CFG, CFG colour, UOXF, UOXF colour and text only. GlycanBuilder can also calculate the masses of glycan structures according to the options selected by the user. GlycanBuilder is a versatile tool for building carbohydrates, with multiple options for exporting the generated structures in the form of text format (GlycoCT, LINUCS, Glycominds, Glyde II) or image (.svg, .png, .jpg, .bmp, .pdf, etc.) files.

DrawRINGS. DrawRINGS [17] is a Java-based applet for rendering glycan structures on canvas (http://www.rings.t.soka.ac.jp/drawRINGS-js/). The different drawing features in an interactive interface (Figure 6, bottom) can be selected with the mouse by surfing the buttons and scroll-down menus. Alternatively, KCF files or KCF text format can be used as input. The free movement of the monosaccharides allows drawings with flexible geometry, for example, for schematic studies of carbohydrates. The drawn glycan structure can be exported in the KCF or IUPAC text format or saved in .png format. The drawn structure can further be used as a query for the search in glycan databases; using match percentage (Similarity) or by the number of components matched (Matched) criteria. Four predefined score matrices are available, named: N-glycans, O-glycans, Sphingolipids and Link_similarity. The "Link_similarity" matrix is based on glycosidic linkages and monosaccharides that may be more highly substituted with other glycosidic linkages and monosaccharides, respectively. There is a query to search the generated structure in the RINGS or GlycomeDB databases (or both). The former compiles data from the KEGG GLYCAN and GLYCOSCIENCES.de databases. DrawRINGS is an efficient tool for sketching glycan figures as well as translating to (and from) the KCF and IUPAC text formats.

DrawGlycan-SNFG. DrawGlycan-SNFG [31] is an open-source program available with a web interface (Figure 7, top) at

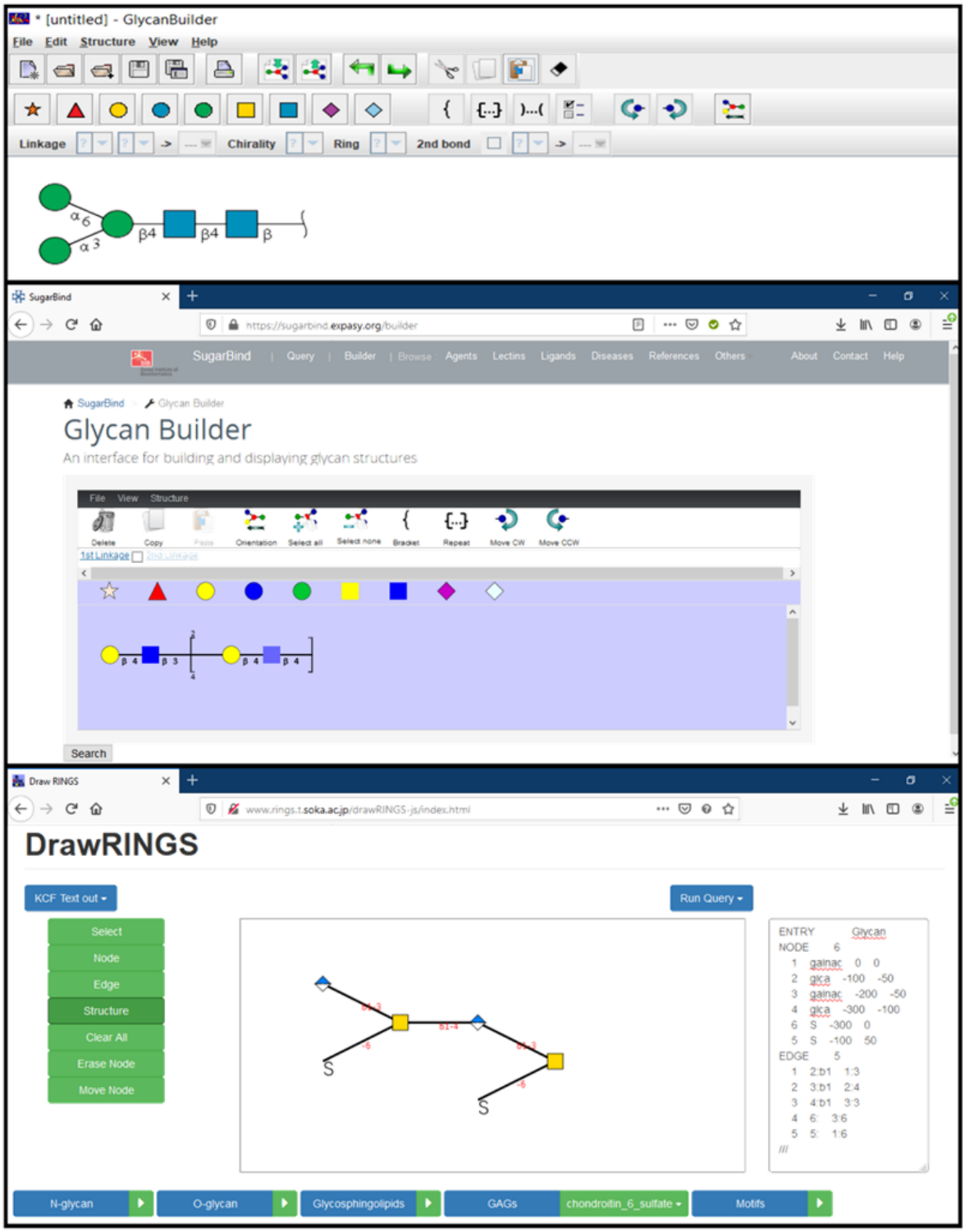


Figure 6: From top to bottom: GlycanBuilder2 [46] interface with a glycan image in SNFG notation. Original GlycanBuilder [47] interface with some of the available templates rendered as images. DrawRINGS [48] interface featuring a glycan and its KCF text output.

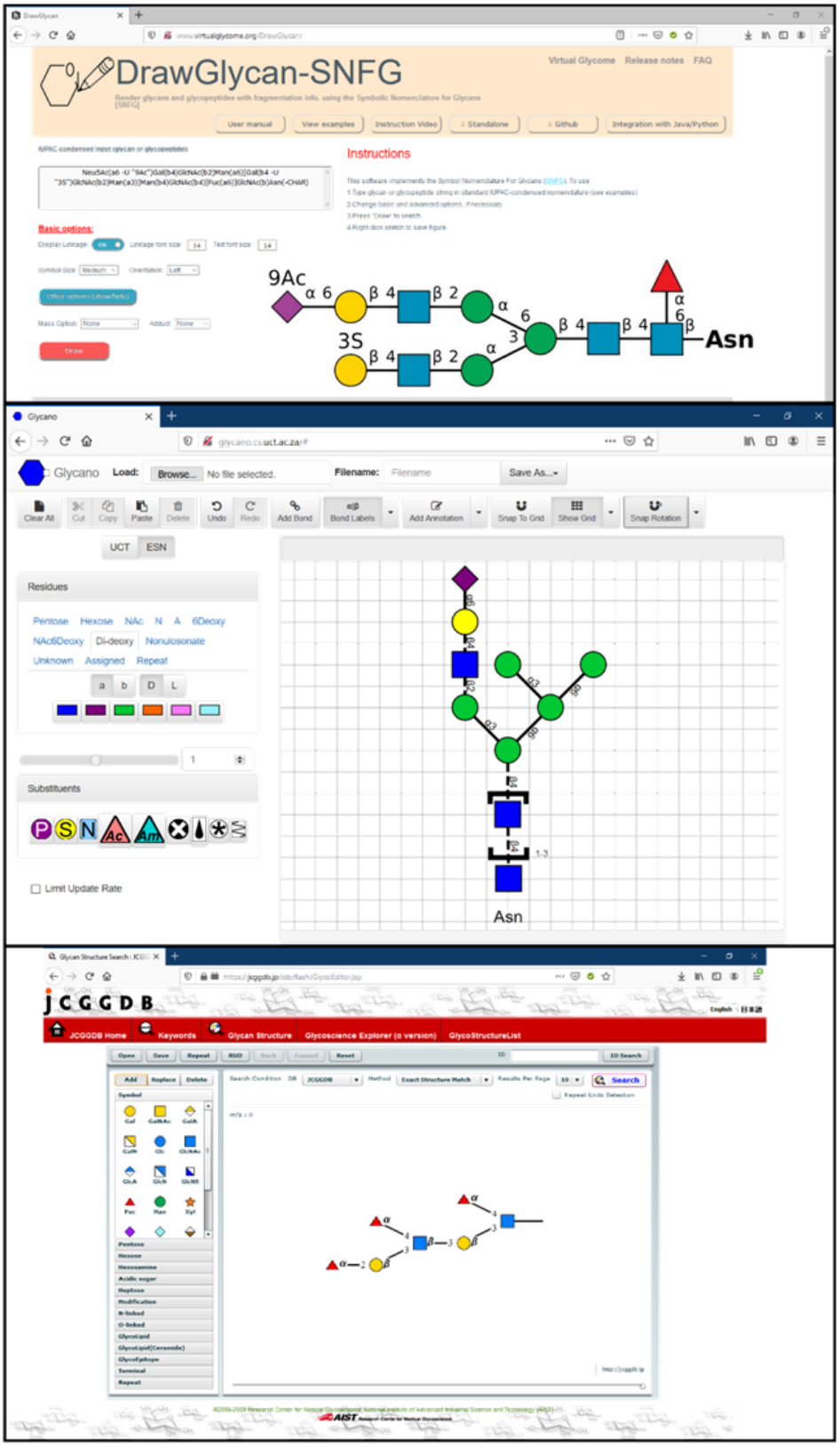


Figure 7: From top to bottom: DrawGlycan-SNFG [51] web interface with a glycan text input and the resulting image output. Glycano [52] interface with a glycan structure. GlycoEditor [53] interface, linkage selection is triggered by adding a new monosaccharide.

http://www.virtualglycome.org/DrawGlycan. The same web page gives access to a downloadable, standalone Graphical User Interface (GUI) version of this tool with additional functionality. It can be launched from different platforms including Windows, Mac or Linux. The program can be used to render glycans and glycopeptides using SNFG and uses IUPAC-condensed text inputs. The DrawGlycan-SNFG version with command-line operations makes it more versatile as it allows integration of multiple features of the program using custom scripts. The tool uses automatic operations for the majority of the drawing, which could meet the needs of researchers, but additional intervention may sometimes be required to get the desired output. For example, manual input in IUPAC-condensed language allows to generate, among others: repeating units, adducts, tethering to other structures (represented by text), and complex branching (the examples section showcases these options). The drawn glycan structure can be saved as .jpg image and modified through parameters such as symbol and text size, the thickness of lines, orientation of drawing and spacing. This software provides all the guidance and tools needed to generate high-quality pictures. DrawGlycan-SNFG is a reliable choice for building glycans.

In addition to glycan structure drawing, DrawGlycan-SNFG (version 2) [54] is equipped with a wide range of options to enhance the usability of the original code [32]. The new version is capable to accommodate the latest updates to the SNFG [7]. This tool has been particularly upgraded for MS spectrum annotation by adding an intuitive interface with additional features. The upgraded version can depict bond fragmentation, repeating structural unit anomeric groups, adduct ions, different types of glycosidic linkages etc. These advanced features make this tool ideal for integrated use with various glycoinformatics software and also for applications in glycoproteomics, glycomics and mass spectrometry (MS). One of the illustrations involves combined use with the gpAnnotate application, dedicated to score and annotate MS/MS glycopeptide spectrums in different fragmentation modes [54].

Glycano. Glycano (available at http://glycano.cs.uct.ac.za) is a software tool for drawing glycans. This tool is based on JavaScript, which can be used without the requirement of any server or browser dependency. The interactive interface allows sketching via the drag-and-drop method on canvas (with or without grid). The software is provided with "UCT" and "ESN", interchangeable interfaces (Figure 7, middle) with different symbols for monosaccharides. These names (UCT and ESN) correspond to the University of Cape Town, South Africa, where Glycano was developed, and to the "Essentials of Glycobiology Symbol Nomenclature", precursor of the SNFG symbol set [55]. The interface provides a wide choice of monosaccharides and substituents represented in SNFG symbols but lacks the standard colour scheme. The user can easily modify the structure with by click and drag, which allows to either cut/copy, delete or move a portion of the structure. The drawn structure can be saved in text format, in .gly format or as an image (PNG and SVG formats). A drawback to note is that linking the monosaccharides at specific positions is only possible in the UCT mode, which means that back-and-forth between the two symbol systems is necessary to define the linkages correctly. Despite some drawbacks, this is an excellent tool due to its ease-of-use, tenable degree of freedom, and functionalities/options for sketching and building glycan structures.

GlycoEditor. GlycoEditor [19] (available at https://jcggdb.jp/idb/flash/GlycoEditor.jsp) is an online software for drawing glycans. Through a straightforward interface, three ways of input are possible: by JCGGDB ID, through a library of common oligosaccharides and by direct input. A list of most common monosaccharides is presented, and the rest can be found categorised by family. The click and drag addition of new monosaccharides trigger the selection of linkage-type and configuration (Figure 7, bottom). The tool provides an option to create repeating units. Additionally, several functionalisation options are also available. Once the structure is ready, the user can save it as an .xml file. GlycoEditor allows searching a given structure across many databases in four ways: exact structure match (with or without anomer and linkage specifics) and the same for substructure match. The central database featured is the JCGGDB, to which can be added, among others: Glaxy, GlycomeDB, GlycoEpitope, GMDB, KEGG, etc. Searching by ID is also possible. GlycoEditor is a now dated tool that allows efficiently building glycans and performing databases searches.

GLYCO.ME (SugarBuilder). Glyco.me-SugarBuilder (available at https://beta.glyco.me/sugarbuilder) is online software for drawing glycans. The interface leads to rapid carbohydrate construction. A panel of monosaccharide templates complements the drawing interface (one pre-built oligosaccharide is available (Figure 8, top). The user can start a chain from amino acid residues: Asn, Ser or Thr, then structure building is limited by to a set of "rules" (limiting building options to known carbohydrates). These rules may be deactivated with a switch button to draw freely. A list of 13 monosaccharides is deployed, and sequential clicking allows their addition to the existing structure and definition of the associated glycosidic bond (the relative sizes of the options available related to their real statistical value for that particular linkage). Upon building some specific motifs, if they are recognised, an option for repeating units appears. Other switch buttons allow the user to change the orientation of the drawing, show/hide linkage information etc. The Oxford notation can be enabled for glycosidic bonds only. The

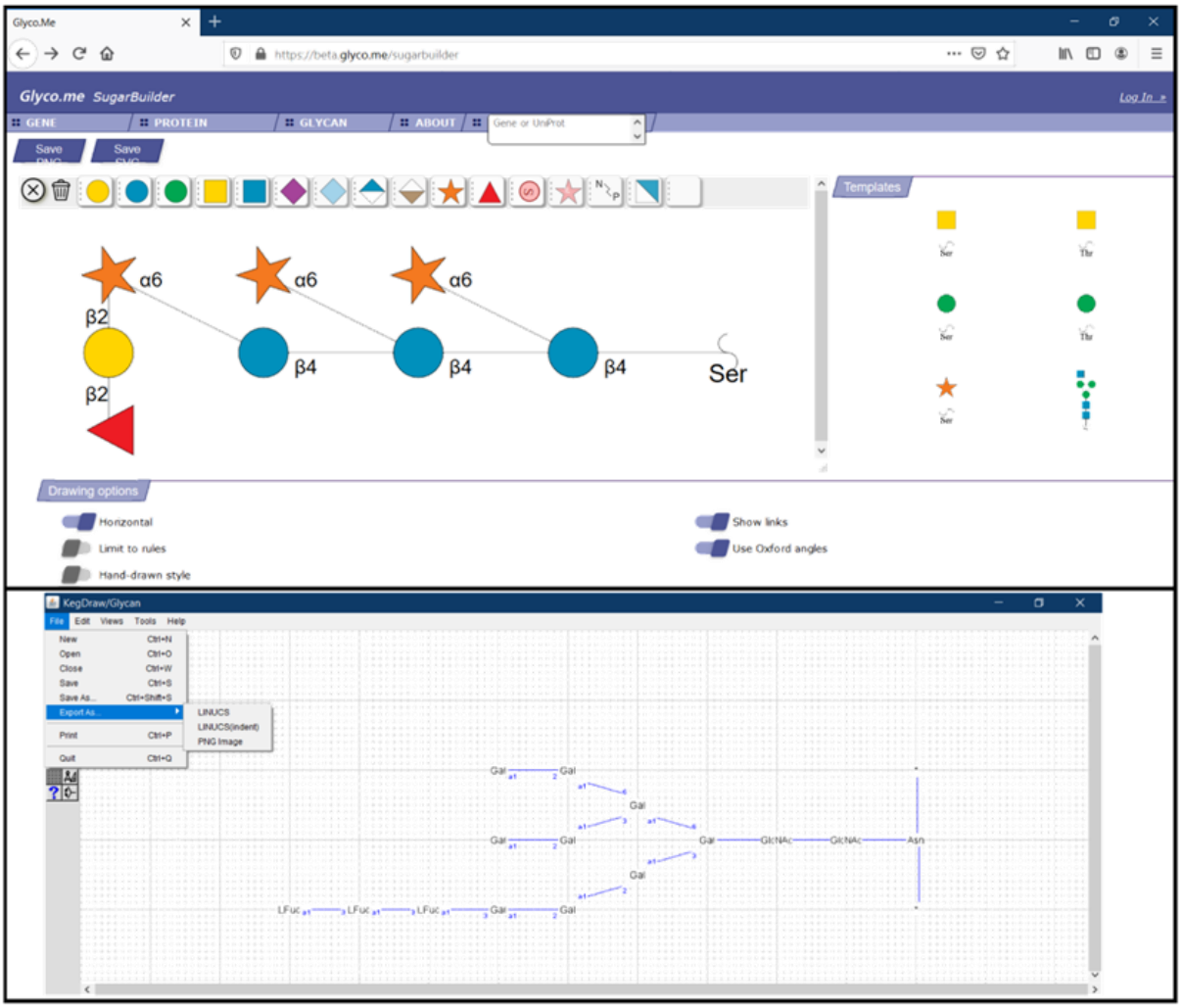


Figure 8: From top to bottom: Glyco.me SugarBuilder [56] interface with a glycan structure showing options to define anomericity and monosaccharide linkage position. KegDraw [57] interface with a glycan structure and available options to save the structure file in different formats.

structure obtained can be rendered as .png or .svg images. Glyco.me-SugarBuilder is still under development: more monosaccharides/substitutions/templates will complete an already very functional platform. The quick and easy options put forward offer natural building and liberty for tailoring the rendered image.

KegDraw. KegDraw (https://www.kegg.jp/kegg/download/kegtools.html) is a freely available Java application for rendering glycan structures. It can be downloaded and installed locally as a platform-independent tool. This tool can be used in two different modes: "Compound mode" which can be used for drawing small molecules (similarly to any chemical structure drawing software), and "Glycan mode" which is dedicated for rendering glycan structures using different monosaccharide units. The simplest method for drawing involves a selection of monosaccharides and glycosidic linkages from an available list to generate a glycan structure. Alternatively, a text box option provides a way to draw uncommon types of monosaccharides. The tool also contains templates from KEGG GLYCAN and their importation using their accession number. Besides, input files in KCF can be used while the output can be saved in LINUCS, KCF or an image in PNG format (Figure 8, bottom). The glycan structure in text format can be further used as a query for search in KEGG GLYCAN and CarbBank databases. Hence, KegDraw can be an option for the freely available tool for drawing and querying chemical structures. However, there are similar tools already available for glycan drawing with more advanced and acceptable notations.

Glycan builders

Sweet II. Sweet [58] is a web-based program for constructing 3D models of glycans from a sequence using standard nomenclature accessible at http://www.glycosciences.de/modeling/

sweet2/doc/index.php (Figure 9, top). This tool is available as a part of the glycosciences.de website, which also provides other options for analysing glycans in three-dimensional space. This program uses a glycan sequence in a standard format and generates a 3D model in the form of a .pdb file. The glycan input can come from a library of relevant oligosaccharides, available through one of the sub-menus. Alternatively, manual input is possible in three platforms adapted for increasing complexity. The model can be further minimised using MM2 [59] and MM3 [60] methods. The 3D models can be viewed using molecular viewers like JMol, WebMol-applet, Chemis3D-applet, etc. Besides, the program also generates additional files which can be used for molecular mechanics and molecular dynamics using molecular modelling tool like Tinker [61]. This tool is as a versatile tool for generating a 3D model for glycans.

GLYCAM-web (Carbohydrate Builder). Carbohydrate builder [65] is an online tool (at http://glycam.org/) for carbohydrate structure drawing and subsequent 3D structure building. With a flexible interface, it uses three methods for glycan building. The first method is manual building ("Carbohydrate Builder" button). It allows selection of monosaccharide, as well as defining linkages, branching and substitution (Figure 9, middle). The second method involves the use of a template library (using "Oligosaccharide libraries" button) containing commonly relevant structures (http://glycam.org/Pre-builtLibraries.jsp). The third option (direct input from a text sequence) becomes relevant when the glycan structure does not exist in the library or challenging to build due to structural complexity. In this case, a text for the oligosaccharide in GLYCAM-Web's condensed notation can be entered as an input to create the glycan structure. Once the glycan is generated, the options include the solvation of the structure and the manual input of the glycosidic linkages. The tool allows structure minimisation and generates rotamers which can be visualised using JSmol viewer. Information about the force field that is used to build the structure is also provided. The multiple structures can be downloaded compressed as .tar, .gz or .zip files containing .pdb files. Similarly, the 2D image can be saved in GIF format. GLYCAM-web- Carbohydrate Builder can be used to prepare the system for MD simulation as it solvates the glycans and also generates the topology and coordinate files. In addition to its carbohydrate builder, Glycam-web consists of additional tools like glycoprotein builder and glycosaminoglycans (GAG) builder.

CHARMM-GUI (Glycan Reader and Modeler). The CHARMM-GUI (http://www.charmm-gui.org) is a web-based graphical user interface which provides various functional modules to prepare complex biomolecular systems and input files for molecular simulations. Glycan Reader and Modeler [65-67] is a part of CHARMM-GUI (Figure 9, bottom) and available as a freely accessible online tool at http://charmm-gui.org/input/glycan. It can read input files in PDB, PDBx/mmCIF and CHARMM formats containing glycans and automatically detects the carbohydrate molecules and glycosidic linkage information. Alternatively, it can also read a glycan sequence (GRS format) to generate a 3D model and input files for MD simulation of the carbohydrate-only system. GRS carbohydrate sequences can be made through a straightforward interface: monosaccharides (20+ options) and their linkages are added incrementally from drop-down menus. A useful feature of this tool is the real-time rendering of the carbohydrate image: each added monosaccharide and modified linkage is directly reported to the image as well as to a text (GRS) format. Option for numerous chemical modifications is also available.

On the other hand, the Glycan Modeler allows in silico N-/O-glycosylation for glycan-protein complexes and generates a "most relevant" glycan structure through Glycan Fragment Database (GFDB) [68] search which gives proper orientations relative to the target protein. In the absence of target glycan sequence in GFDB, the structures are generated by using the valid internal coordinate information (averaged phi, psi, and omega glycosidic torsion angles) in the CHARMM force field. Input files for CHARMM can be generated for the purpose of MD simulation. Amongst other possible outputs, 3D representations of the glycans are available as .pdb files. This tool can be helpful for researchers to generate 2D depictions of a glycan and then obtain the corresponding 3D representation, which can be useful for modelling studies of glycans and glycoconjugates.

doGlycans. doGlycans [69] is a compilation of tools dedicated for preparing carbohydrate structures for atomistic simulations of glycoproteins, carbohydrate polymers and glycolipids using GROMACS [70,71] In the form of Python scripts; the tools are used to prepare the system, which generally includes the processing of a.pdb file using the *pdb2gmx* tool. Subsequently, a glycosylation model can be prepared for carbohydrate polymer simulation using the *prepreader.py* script. Similarly, the *doglycans.py* script can be used to develop models for glycoproteins and glycolipids. Together, these tools are called doGlycans toolset. Although doGlycans is highly flexible, it only uses the sugar units that are defined in GLYCAM. The topologies generated for glycosylated proteins and glycolipids are compatible with the OPLS [72] and AMBER [73] force fields. The topology for carbohydrate polymers is based on the GLYCAM force field. The user needs to provide the ceramide topology as input to generate the topologies for glycolipids. The tools contained in doGlycans create 3D models and simulation files as a starting point for more complex molecular simulation studies.

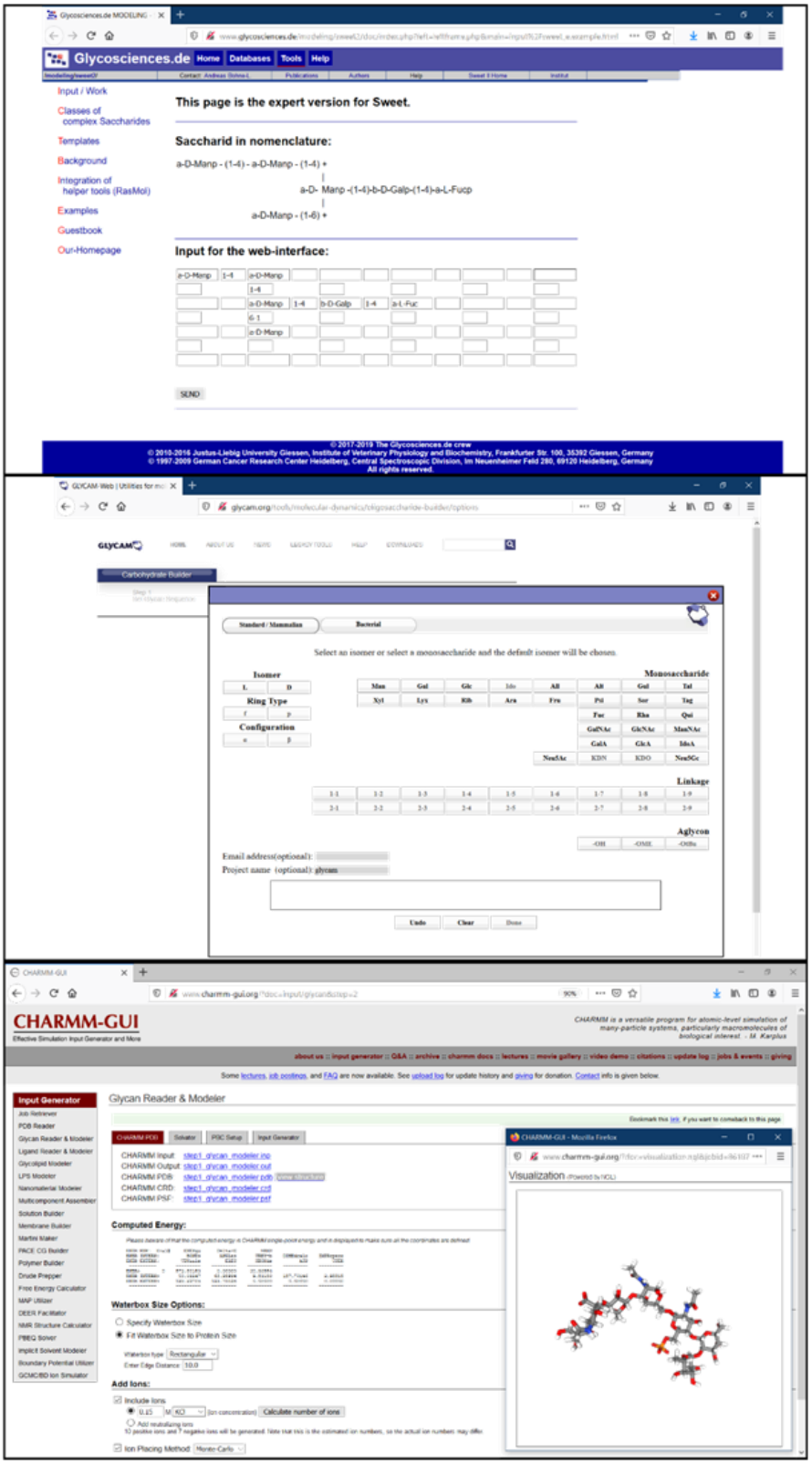


Figure 9: From top to bottom: Sweet II [62] web-interface with a text input to generate a 3D model. GLYCAM Carbohydrate Builder [63] interface which accepts a text input for glycans and generates 3D models. CHARMM-GUI (Glycan reader and Modeler) [64] interface with a 3D structure output generated using a glycan sequence as input.

RosettaCarbohydrate. Rosetta is a software suite for macromolecular modelling as an extensive collection of computer code mostly written in C++ and Python languages. Rosetta is available to academic and commercial researchers through a license available at https://www.rosettacommons.org/software/license-and-download. The licence is free for academic users. The tool runs best on Linux or macOS platforms only. It can be installed on a multiprocessor computing cluster to increase efficiency. RosettaCarbohydrate [74,75] tool provides the methods for general modelling and docking applications for glycans and glycoconjugates. The application accepts the standard PDB, GLYCAM, and GlycoWorkbench (.gws) file formats and the available utilities (codes) helps with the general problems in sampling, scoring, and nomenclature related to glycan modelling. It samples glycosidic bonds, ring forms, side-chain conformations, and utilises a glycan-specific term within its scoring function. The tool also consists of utilities for virtual glycosylation, protein–glyco-ligand docking, and glycan "loop" modelling. This tool is best for the researcher with basic knowledge and skills to work with a command-line interface (Linux).

PolysGlycanBuilder. PolysGlycanBuilder [76] is a web-based tool (http://glycan-builder.cermav.cnrs.fr/) with an interactive and more usable interface (Figure 10). The software translates a glycan sequence or polysaccharide repeat unit into the coordinate set of the corresponding tertiary structure, in one or several of its low energy conformations. The construction follows an intuitive scheme which is as close as possible to the way glycoscientists draw the sequence of their structures. The simplest method for model building involves dragging and dropping monosaccharide units to the canvas or workspace grid. The software displays rows of monosaccharides in the form of standard SNFG symbols with 3D information (furanose/pyranose shape, configuration, anomericity, and ring conformation). Glycosidic linkages can be easily defined, as the values of the dihedral angles (Φ, Ψ, Ω). They can be manually set or extracted from a database of low energy conformations of 600 disaccharide segments. The monosaccharides have been subjected to geometry optimisation using molecular mechanics approach. For a given input sequence, the corresponding 3D coordinates are generated at the PDB format. Within the process of construction, the structure is displayed via the LiteMol and eventually optimised to remove any steric clashes. The image for the glycan can be downloaded and saved in SVG format. Keeping the glycan/polysaccharide structure in text format (condensed IUAPC, GlycoCT, SNFG and INP) offers several ways to connect to other applications. Other than drag and drop method, Polys-Glycan-Builder also accepts input of files in INP, IUPAC and GlycoCT formats. An interactive interface accompanies the application, which makes it more versatile for glycan drawing and 3D model building.

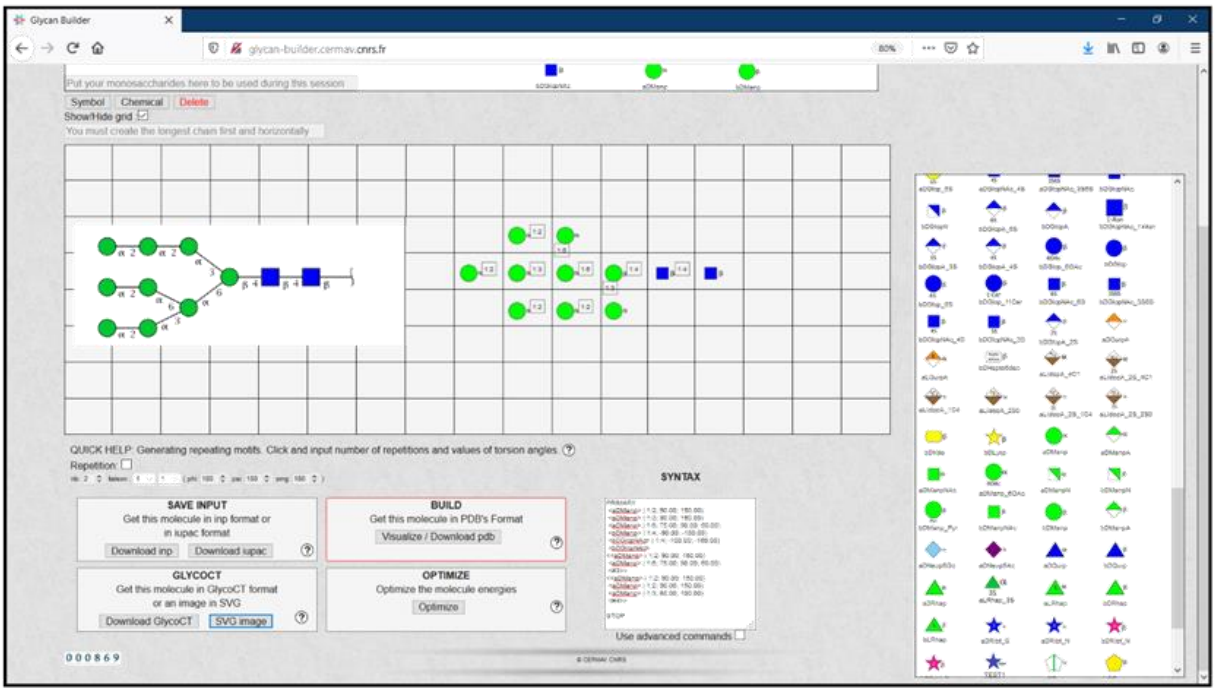


Figure 10: PolysGlycanBuilder [77] interface illustrating glycan drawing using SNFG symbols. The glycan can be further converted into a 3D model.

Displaying 3D structures of glycans

3D-SNFG VMD interface and visualisation algorithms. The recently introduced 3D-Symbol Nomenclature for Glycans (3D-SNFG) [15] allows the representation of carbohydrates in an unusual way: the SNFG symbols are added to a three-dimensional structure. The 3D-SNFG script must be integrated into the visual molecular dynamics (VMD) [21,78] viewer software to enable the representation of glycans as large SNFG-matching3D shapes that can either replace the molecular monosaccharides or stay lodged at the geometric centre of the cycle (Figure 11, top left). Upon the input of a glycan-containing structure (in PDB format), the integrated script in VMD automatically recognises the common monosaccharide names and generates the 3D shapes. The embedded script also enables shortcuts keys from keyboard to quickly change between large and small 3D-SNFG shapes and also label the reducing terminus. The 3D structure displayed in VMD can be saved as a .bmp image file. Thanks to 3D-SNFG, the standardised representation of glycan structures can finally take a step into the 3D space. The obtained images can become very useful for quick assessment of 3D glycan models.

In addition to the 3D-SNFG script, *PaperChain* and *Twister* [83] are two visualisation algorithms available with the Visual Molecular Dynamics (VMD) package. These algorithms are useful to visualize complex cyclic molecules and multi-branched polysaccharides.{Cross, 2009 #69} *PaperChain* displays rings in a molecular structure with a polygon and colours them according to the ring pucker. The other algorithm (*Twister)* traces glycosidic bonds in a ribbon representation that twists and changes its orientation according to the relative position of following sugar residues, hence provides an important conformational detail in polysaccharides. Combination of these algorithms with other visualisation features available in VMD can enhance the flexibility of displaying structural details of glycoconjugate, glycoprotein and cyclic structures.

LiteMol. The LiteMol [22] viewer is a freely available web application (Figure 11, top right) for 3D visualisation of macromolecules and other related data. LiteMol enables standard visualisation of macromolecules in different representation modes like surface, cartoons, ball-and-stick, etc. The software can be accessed at v.litemol.org and also available for integration in a webpage from the github (https://github.com/dsehnal/LiteMol). LiteMol is compatible with all modern browsers without the support of additional plugins. The viewer automatically depicts any carbohydrate residues and displays 3D structures of carbohydrates with 3D-SNFG symbols, which allows the viewer to identify the monosaccharides readily. The presented structure can be saved as a.png image file. Any monosaccharide with a residue name in PDB can be visualised using 3D-SNFG in LiteMol. However, a significant portion of the carbohydrates may contain some form of error in annotation, which would result in either no symbol or an incorrect symbol. Although LiteMol is an efficient and rapid 3D viewer for glycans, 3D representation does not provide any information about the glycosidic linkage type (e.g. α1-3 or β1-4). Also, it does not display any information about connection and configuration. If this information is required, returning to the classic molecular representation is possible.

PyMOL- Azahar plugin. Azahar [84] is a plugin in PyMOL [85] which enables building, visualization and analysis of glycans and glycoconjugates. This tool is based on Python and provides additional computing environment within the PyMOL package. The tool is provided with a template list of saccharide structures to facilitate structure building and visualisation. The interface provides three option menus to assist glycan structure building. The two first options help to specify residues to be connected from a list of available templates, and the third one allows selection of the chemical bond between the residues. The visualisation using PyMOL includes three cartoon-like representations. These display modes provided in the tool simplify the representation of glycan structures in cartoon, wire and bead representations. In cartoon and wire representations, the rings in sugars are shown as non-flat polygons connected by rods while in the bead representation mode, these cycles are represented as a sphere. In addition of visualization of static structures, the tool also allows analysis of trajectories of MD simulations. The tool can be used for conformational search using a Monte Carlo approach [86]. The conformational search is done by perturbing a torsional angle, followed by an energy minimization using the MMFF94 force field. Azahar is freely accessible from http://www.pymolwiki.org/index.php/Azahar.

UnityMol/SweetUnityMol. Sweet UnityMol [32] is a molecular structure viewer (Figure 11, middle) developed from the game engine Unity3D. The software is available for free download (https://sourceforge.net/projects/unitymol/files/UnityMol_1.0.37/) from the SourceForge project website. It can be installed in Mac, Windows and Linux platforms. The program reads files in PDB, mmCIF, Mol2, GRO, XYZ, and SDF formats, OpenDX potential maps and XTC trajectory files. It efficiently displays specific structural features for the simplest to the most complex carbohydrate-containing biomolecules. Sweet UnityMol displays 3D carbohydrate structures with different modes of representation, such as: liquorice, ball-and-stick, hyperBalls, RingBlending, hydrophilic/hydrophobic character of sugar face etc. The most recent version is fully compatible with the SNFG colour coding, which also uses acceptable pictorial representation, generally used in carbohydrate chemistry, biochemistry and glycobiology.

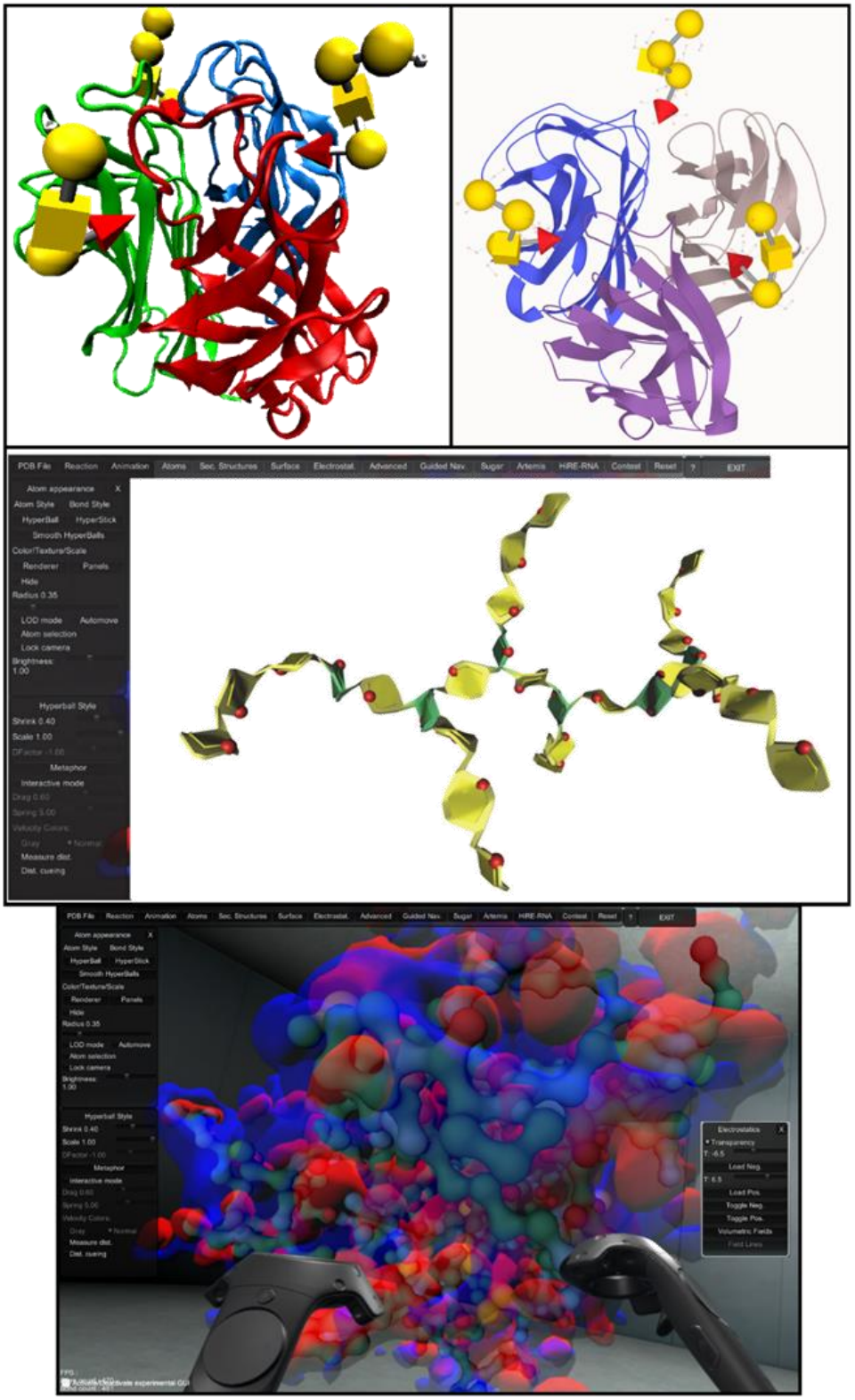


Figure 11: From top to bottom: 3D-SNFG representation of glycan using 3D-SNFG script integrated VMD [79]. LiteMol [80] interface with 3D-SNFG representation of glycan in a protein–glycan complex. SweetUnityMol [81] among the several types of representations a ribbon-like display of polysaccharide ribbons maintains the SNFG colour coding of monosaccharides. UnityMol [82] within an immersive virtual reality context.

SweetUnityMol provides a continuum from the conventional ways to depict the primary structures of complex carbohydrates all the way to visualising their 3D structures. Several options are offered to the user to select the most relevant type of depictions, including new features, such as "Coarse-Grain" representation while keeping the option to display the details of the atomic representations. Powerful rendering methods produce high-quality images of molecular structures, bio-macromolecular surfaces and molecular interactions.

A recently developed version of UnityMol has been implemented with the immersive Virtual Reality context using head-mounted displays [87]. It offers high-quality visual representations, ease of interactions with multiple molecular objects, powerful tools for visual manipulations, accompanied by the evaluation of intermolecular interactions. Consequently, simultaneous investigations of multiple objects such as macromolecular interactions gain in efficiency and accuracy. (Figure 11, bottom).

Conclusion

The set of computational tools presented above illustrates the rich contributions of a community devoted to enabling the accurate representation of complex carbohydrates via the development and implementation of a versatile informatics toolbox. These legitimate efforts aim at facilitating communication within the scientific community. To establish a comparative analysis of the several available applications, we evaluated 17 selected items that characterise best their availability, implementation, maintenance and field of use. The comparative analysis of tools could be useful for glycobiologists or any researcher looking for a ready to use, simple application for the sketching, building and display of glycans.

This article provides an overview of the computational tools and resources available for glycan sketching, building and representing. It also provides a descriptive analysis of the recently developed software tools dedicated explicitly to glycans and glycoconjugates. The newly developed tools are more advanced and use the standard nomenclature and symbols for glycan representation. These tools can further help to standardise the description of glycans in research, communication and databases.

Supporting Information

Supporting Information File 1

Features of glycan sketchers, builders and viewers.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-16-199-S1.pdf]

Acknowledgements

Appreciation is extended to Drs. A. Imberty, A. Varrot, L. Belvisi and A. Bernardi for their support.

Funding

This research was performed within the framework of the PhD4GlycoDrug Innovative Training Network and was funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 765581. The work was supported by the Cross-Disciplinary Program Glyco@Alps, within the framework "Investissement d'Avenir" program [ANR-15IDEX-02].

ORCID® iDs

Kanhaya Lal - https://orcid.org/0000-0001-8555-7948
Rafael Bermeo - https://orcid.org/0000-0002-4451-878X
Serge Perez - https://orcid.org/0000-0003-3464-5352

References

1. Alocci, D.; Lisacek, F.; Perez, S. *A Traveler's Guide to Complex Carbohydrates in the Cyber Space*. http://www.glycopedia.eu/IMG/pdf/traveler_s_guide_to_cyber_space.pdf
2. Perez, S.; Aoki-Kinoshita, K. F. *Development of Carbohydrate Nomenclature and Representation;* Springer, 2017; pp 7–25. doi:10.1007/978-4-431-56454-6_2
3. Kornfeld, S.; Li, E.; Tabas, I. *J. Biol. Chem.* **1978,** *253,* 7771–7778.
4. Royle, L.; Dwek, R. A.; Rudd, P. M. *Curr. Protoc. Protein Sci.* **2006,** *43,* 12.6.1–12.6.45. doi:10.1002/0471140864.ps1206s43
5. Harvey, D. J.; Merry, A. H.; Royle, L.; Campbell, M. P.; Dwek, R. A.; Rudd, P. M. *Proteomics* **2009,** *9,* 3796–8301. doi:10.1002/pmic.200900096
6. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Marth, J. D.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E. *Proteomics* **2009,** *9,* 5398–5399. doi:10.1002/pmic.200900708
7. Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lütteke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; Varki, A.; Woods, R. J.; Darvill, A.; Dell, A.; Henrissat, B.; Bertozzi, C.; Hart, G.; Narimatsu, H.; Freeze, H.; Yamada, I.; Paulson, J.; Prestegard, J.; Marth, J.; Vliegenthart, J. F. G.; Etzler, M.; Aebi, M.; Kanehisa, M.; Taniguchi, N.; Edwards, N.; Rudd, P.; Seeberger, P.; Mazumder, R.; Ranzinger, R.; Cummings, R.; Schnaar, R.; Perez, S.; Kornfeld, S.; Kinoshita, T.; York, W.; Knirel, Y. *Glycobiology* **2019,** *29,* 620–624. doi:10.1093/glycob/cwz045
8. Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T.; Prestegard, J. J.; Schnaar, R. L.; Freeze, H. H.; Marth, J. D.; Bertozzi, C. R.; Etzler, M. E.; Frank, M.; Vliegenthart, J. F. G.; Lütteke, T.; Perez, S.; Bolton, E.; Rudd, P.; Paulson, J.; Kanehisa, M.; Toukach, P.; Aoki-Kinoshita, K. F.; Dell, A.; Narimatsu, H.; York, W.; Taniguchi, N.; Kornfeld, S. *Glycobiology* **2015,** *25,* 1323–1324. doi:10.1093/glycob/cwv091
9. Aoki, K. F.; Yamaguchi, A.; Ueda, N.; Akutsu, T.; Mamitsuka, H.; Goto, S.; Kanehisa, M. *Nucleic Acids Res.* **2004,** *32,* W267–W272. doi:10.1093/nar/gkh473

10. Ceroni, A.; Dell, A.; Haslam, S. M. *Source Code Biol. Med.* **2007,** *2,* No. 3. doi:10.1186/1751-0473-2-3
11. Damerell, D.; Ceroni, A.; Maass, K.; Ranzinger, R.; Dell, A.; Haslam, S. M. Annotation of Glycomics MS and MS/MS Spectra Using the GlycoWorkbench Software Tool. In *Glycoinformatics. Methods in Molecular Biology;* Lütteke, T.; Frank, M., Eds.; Humana Press: New York, NY, 2015; Vol. 1273, pp 3–15. doi:10.1007/978-1-4939-2343-4_1
12. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000,** *28,* 235–242. doi:10.1093/nar/28.1.235
13. Engelsen, S. B.; Hansen, P. I.; Pérez, S. *Biopolymers* **2014,** *101,* 733–743. doi:10.1002/bip.22449
14. Alocci, D.; Suchánková, P.; Costa, R.; Hory, N.; Mariethoz, J.; Vařeková, R.; Toukach, P.; Lisacek, F. *Molecules* **2018,** *23,* 3206. doi:10.3390/molecules23123206
15. Thieker, D. F.; Hadden, J. A.; Schulten, K.; Woods, R. J. *Glycobiology* **2016,** *26,* 786–787. doi:10.1093/glycob/cww076
16. McNaught, A. D. *Adv. Carbohydr. Chem. Biochem.* **1997,** *52,* 44–177. doi:10.1016/s0065-2318(08)60090-6
17. Akune, Y.; Hosoda, M.; Kaiya, S.; Shinmachi, D.; Aoki-Kinoshita, K. F. *OMICS* **2010,** *14,* 475–486. doi:10.1089/omi.2009.0129
18. Hashimoto, K.; Goto, S.; Kawano, S.; Aoki-Kinoshita, K. F.; Ueda, N.; Hamajima, M.; Kawasaki, T.; Kanehisa, M. *Glycobiology* **2006,** *16,* 63R–70R. doi:10.1093/glycob/cwj010
19. Maeda, M.; Fujita, N.; Suzuki,, Y.; Sawaki, H.; Shikanai, T.; Narimatsu, H. JCGGDB: Japan Consortium for Glycobiology and Glycotechnology Database. In *Glycoinformatics. Methods in Molecular Biology;* Lütteke, T.; Frank, M., Eds.; Humana Press: New York, NY, 2015; Vol. 1273, pp 161–179. doi:10.1007/978-1-4939-2343-4_12
20. Damerell, D.; Ceroni, A.; Maass, K.; Ranzinger, R.; Dell, A.; Haslam, S. M. *Biol. Chem.* **2012,** *393,* 1357–1362. doi:10.1515/hsz-2012-0135
21. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996,** *14,* 33–38. doi:10.1016/0263-7855(96)00018-5
22. Sehnal, D.; Grant, O. C. *J. Proteome Res.* **2019,** *18,* 770–774. doi:10.1021/acs.jproteome.8b00473
23. Foley, B. L.; Tessier, M. B.; Woods, R. J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2,* 652–697. doi:10.1002/wcms.89
24. Mallajosyula, S. S.; Guvench, O.; Hatcher, E.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2012,** *8,* 759–776. doi:10.1021/ct200792v
25. Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2009,** *5,* 2353–2370. doi:10.1021/ct900242e
26. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008,** *29,* 622–655. doi:10.1002/jcc.20820
27. Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005,** *26,* 1400–1412. doi:10.1002/jcc.20275
28. Molinero, V.; Goddard, W. A. *J. Phys. Chem. B* **2004,** *108,* 1414–1427. doi:10.1021/jp0354752
29. Herget, S.; Ranzinger, R.; Maass, K.; Lieth, C.-W. v. d. *Carbohydr. Res.* **2008,** *343,* 2162–2171. doi:10.1016/j.carres.2008.03.011
30. Tanaka, K.; Aoki-Kinoshita, K. F.; Kotera, M.; Sawaki, H.; Tsuchiya, S.; Fujita, N.; Shikanai, T.; Kato, M.; Kawano, S.; Yamada, I.; Narimatsu, H. *J. Chem. Inf. Model.* **2014,** *54,* 1558–1566. doi:10.1021/ci400571e
31. Cheng, K.; Zhou, Y.; Neelamegham, S. *Glycobiology* **2017,** *27,* 200–205. doi:10.1093/glycob/cww115
32. Perez, S.; Tubiana, T.; Imberty, A.; Baaden, M. *Glycobiology* **2015,** *25,* 483–491. doi:10.1093/glycob/cwu133
33. Yamada, I.; Shiota, M.; Shinmachi, D.; Ono, T.; Tsuchiya, S.; Hosoda, M.; Fujita, A.; Aoki, N. P.; Watanabe, Y.; Fujita, N.; Angata, K.; Kaji, H.; Narimatsu, H.; Okuda, S.; Aoki-Kinoshita, K. F. *Nat. Methods* **2020,** *17,* 649–650. doi:10.1038/s41592-020-0879-8
34. Tiemeyer, M.; Aoki, K.; Paulson, J.; Cummings, R. D.; York, W. S.; Karlsson, N. G.; Lisacek, F.; Packer, N. H.; Campbell, M. P.; Aoki, N. P.; Fujita, A.; Matsubara, M.; Shinmachi, D.; Tsuchiya, S.; Yamada, I.; Pierce, M.; Ranzinger, R.; Narimatsu, H.; Aoki-Kinoshita, K. F. *Glycobiology* **2017,** *27,* 915–919. doi:10.1093/glycob/cwx066
35. Lütteke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C.-W. *Glycobiology* **2006,** *16,* 71R–81R. doi:10.1093/glycob/cwj049
36. *Sugar Sketcher.* https://glycoproteome.expasy.org/sugarsketcher/ (accessed April 2020).
37. *GlycoGlyph.* https://glycotoolkit.com/Tools/GlycoGlyph/ (accessed April 2020).
38. *GlyTouCan.* https://glytoucan.org/ (accessed April 2020).
39. Mehta, A. Y.; Cummings, R. D. *Bioinformatics* **2020,** *36,* 3613–3614. doi:10.1093/bioinformatics/btaa190
40. Tsuchiya, S.; Aoki, N. P.; Shinmachi, D.; Matsubara, M.; Yamada, I.; Aoki-Kinoshita, K. F.; Narimatsu, H. *Carbohydr. Res.* **2017,** *445,* 104–116. doi:10.1016/j.carres.2017.04.015
41. Toukach, P. V.; Egorova, K. S. *Nucleic Acids Res.* **2016,** *44,* D1229–D1236. doi:10.1093/nar/gkv840
42. Kikuchi, N.; Kameyama, A.; Nakaya, S.; Ito, H.; Sato, T.; Shikanai, T.; Takahashi, Y.; Narimatsu, H. *Bioinformatics* **2005,** *21,* 1717–1718. doi:10.1093/bioinformatics/bti152
43. Doubet, S.; Bock, K.; Smith, D.; Darvill, A.; Albersheim, P. *Trends Biochem. Sci.* **1989,** *14,* 475–477. doi:10.1016/0968-0004(89)90175-8
44. Banin, E.; Neuberger, Y.; Altshuler, Y.; Halevi, A.; Inbar, O.; Nir, D.; Dukler, A. *Trends Glycosci. Glycotechnol.* **2002,** *14,* 127–137. doi:10.4052/tigg.14.127
45. Cooper, C. A.; Harrison, M. J.; Wilkins, M. R.; Packer, N. H. *Nucleic Acids Res.* **2001,** *29,* 332–335. doi:10.1093/nar/29.1.332
46. *GlycanBuilder2.* Downloaded from http://www.rings.t.soka.ac.jp/downloads.html (accessed April 2020).
47. *SugarBind GlycanBuilder.* https://sugarbind.expasy.org/builder (accessed April 2020).
48. *DrawRINGS.* http://www.rings.t.soka.ac.jp/drawRINGS-js/ (accessed April 2020).
49. Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M. *J. Proteome Res.* **2008,** *7,* 1650–1659. doi:10.1021/pr7008252
50. Mariethoz, J.; Khatib, K.; Alocci, D.; Campbell, M. P.; Karlsson, N. G.; Packer, N. H.; Mullen, E. H.; Lisacek, F. *Nucleic Acids Res.* **2016,** *44,* D1243–D1250. doi:10.1093/nar/gkv1247
51. *DrawGlycan-SNFG.* http://www.virtualglycome.org/DrawGlycan/ (accessed April 2020).
52. *Glycano.* http://glycano.cs.uct.ac.za/ (accessed April 2020).
53. *GlycoEditor.* https://jcggdb.jp/idb/flash/GlycoEditor.jsp (accessed April 2020).
54. Cheng, K.; Pawlowski, G.; Yu, X.; Zhou, Y.; Neelamegham, S. *Bioinformatics* **2019.** doi:10.1093/bioinformatics/btz819

55. Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., *Essentials of Glycobiology [Internet].* 3 ed.; Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press: 2015–2017.
56. *Glyco.me SugarBuilder.* https://beta.glyco.me/sugarbuilder (accessed April 2020).
57. *KegDraw.* Downloaded from https://www.kegg.jp/kegg/download/kegtools.html (accessed April 2020).
58. Bohne, A.; Lang, E.; von der Lieth, C. W. *Bioinformatics* **1999,** *15,* 767–768. doi:10.1093/bioinformatics/15.9.767
59. Allinger, N. L. *J. Am. Chem. Soc.* **1977,** *99,* 8127–8134. doi:10.1021/ja00467a001
60. Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989,** *111,* 8566–8575. doi:10.1021/ja00205a002
61. Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. *J. Chem. Theory Comput.* **2018,** *14,* 5273–5289. doi:10.1021/acs.jctc.8b00529
62. *Sweet.* http://www.glycosciences.de/modeling/sweet2/doc/index.php (accessed April 2020).
63. *GLYCAM Web.* (2005–2020) Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (http://glycam.org).
64. *CHARMM-GUI Glycan Reader & Modeler.* http://www.charmm-gui.org/?doc=input/glycan (accessed April 2020).
65. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008,** *29,* 1859–1865. doi:10.1002/jcc.20945
66. Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D., Jr.; Im, W. *J. Comput. Chem.* **2011,** *32,* 3135–3141. doi:10.1002/jcc.21886
67. Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. *Glycobiology* **2019,** *29,* 320–331. doi:10.1093/glycob/cwz003
68. Jo, S.; Im, W. *Nucleic Acids Res.* **2013,** *41,* D470–D474. doi:10.1093/nar/gks987
69. Danne, R.; Poojari, C.; Martinez-Seara, H.; Rissanen, S.; Lolicato, F.; Róg, T.; Vattulainen, I. *J. Chem. Inf. Model.* **2017,** *57,* 2401–2406. doi:10.1021/acs.jcim.7b00237
70. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008,** *4,* 435–447. doi:10.1021/ct700301q
71. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005,** *26,* 1701–1718. doi:10.1002/jcc.20291
72. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. *J. Chem. Theory Comput.* **2016,** *12,* 281–296. doi:10.1021/acs.jctc.5b00864
73. *AMBER 2018;* University of California: San Francisco, 2018.
74. Labonte, J. W.; Adolf-Bryfogle, J.; Schief, W. R.; Gray, J. J. *J. Comput. Chem.* **2017,** *38,* 276–287. doi:10.1002/jcc.24679
75. Frenz, B.; Rämisch, S.; Borst, A. J.; Walls, A. C.; Adolf-Bryfogle, J.; Schief, W. R.; Veesler, D.; DiMaio, F. *Structure* **2019,** *27,* 134–139.e3. doi:10.1016/j.str.2018.09.006
76. Perez, S.; Rivet, A., *Methods in Molecular Biology, Glycoinformatics, Methods and Protocols.* 2nd ed.; 2020, (in press).
77. *PolysGlycanBuilder.* http://glycan-builder.cermav.cnrs.fr/ (accessed April 2020).
78. Kuttel, M.; Gain, J.; Burger, A.; Eborn, I. *J. Mol. Graphics Modell.* **2006,** *25,* 380–388. doi:10.1016/j.jmgm.2006.02.007
79. *3D-SNFG.* Downloaded from http://glycam.org/3d-snfg (accessed April 2020).
80. *LiteMol.* https://v.litemol.org/ (accessed April 2020).
81. *SweetUnityMol.* https://sourceforge.net/projects/unitymol/files/OtherVersions/UnityMol-r676-SweetUnityMol/ (accessed April 2020).
82. *UnityMol.* https://sourceforge.net/projects/unitymol/files/ (accessed April 2020).
83. Cross, S.; Kuttel, M. M.; Stone, J. E.; Gain, J. E. *J. Mol. Graphics Modell.* **2009,** *28,* 131–139. doi:10.1016/j.jmgm.2009.04.010
84. Arroyuelo, A.; Vila, J. A.; Martin, O. A. *J. Comput.-Aided Mol. Des.* **2016,** *30,* 619–624. doi:10.1007/s10822-016-9944-x
85. *PyMOL: An open-source molecular graphics tool;* DeLano Scientific, 2002, http://www.pymol.org.
86. Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84,* 6611–6615. doi:10.1073/pnas.84.19.6611
87. Martinez, X.; Chavent, M.; Baaden, M. *Biochem. Soc. Trans.* **2020,** *48,* 499–506. doi:10.1042/bst20190621

License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0). Please note that the reuse, redistribution and reproduction in particular requires that the authors and source are credited.

The license is subject to the *Beilstein Journal of Organic Chemistry* terms and conditions: (https://www.beilstein-journals.org/bjoc)

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.16.199

Table S1. Schematic summary of important features of glycan sketchers, builders and viewers.

S.No.	Software Tool	2D sketchers and builders																					
													3D Builders							3D (representation) viewers			
		Sugar Sketcher	LiGraph	Glyco Glyph	Glycan Builder2 SNFG	Sugarbind Glycan Builder	Draw RingS	Draw Glycan SNFG	Glycano	Glyco Editor	Glyco.me Sugar Builder	KegDraw	Polys Glycan Builder	Sweet	CHARMM Gui	Glycam Carbo Builder	do Glycans	Rosetta Carbohydrate	Carb Builder	3D-SNFG	PyMOL	Sweet UnityMol	LiteMol
1	Available online	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓					✓		✓
2	Download for local installation	✓		✓	✓			✓	✓			✓					✓	✓	✓	✓	✓	✓	✓
3	Instructions		✓	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
4	Self-explanatory interface	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓						✓	
5	Image output (file type)	.SVG	.SVG	.SVG	.PNG, .SVG	.PNG .SVG .BMP ...	.PNG (save as)	.JPG (save as)	.PNG .SVG	(screen shot)	.PNG .SVG	.PNG	.SVG		(screen shot)	.GIF (save as)				.BMP	.PNG	.PNG	.PNG
6	SNFG colours	CMYK	CMYK	RGB	RGB		CMYK	RGB		RGB	RGB		CMYK		CMYK					CMYK	RGB		RGB
7	Oxford linkage geometry	✓			✓	✓	✓	✓	✓		✓		✓										
8	Functionalization	✓		✓	✓	✓		✓	✓	✓		✓			✓	✓	✓	✓			✓		
9	Repeating units [ ]	✓			✓	✓	✓	✓	✓	✓		✓	✓								✓		
10	Linking as glyco-conjugates				✓		✓	✓				✓					✓	✓	✓	✓	✓		✓
11	Text input	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓		✓	✓	✓	✓				✓
12	Option to modify by coding	✓	✓	✓			✓	✓					✓	✓		✓				✓	✓		
13	3D-model visualization												✓	✓	✓	✓				✓	✓		✓
14	3D model output												✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
15	Text/String output (any format)	✓	✓	✓	✓	✓	✓			✓		✓	✓	✓	✓	✓			✓	✓			
16	Glycan library	✓		✓	✓	✓	✓			✓				✓							✓		
17	No. of templates (monosaccharides)	70+	15+	80+	70+	50+	70+	60+	60+	40+	10+	10+	100+	40+	20+	30+				40+			40+

8.2. Prediction and Validation of a Druggable Site on Virulence Factor of Drug Resistant *Burkholderia cenocepacia*

As stated in Sections 3 and 5, this project benefitted from the collaboration of Kanhaya Lal, PhD student from PhD4GlycoDrug. His project 'Design of lectin antagonists through fragment-based screening and molecular modelling', led to the publication of the article presented in this sub-section. It describes the virtual screening of a fragment for a site vicinal to the CRD of BC2L-C-N_{ter}, the biophysical evaluation of fragment hits, and the validation of the virtual structure-based strategy by a new crystal structure featuring a fragment hit in its predicted binding pose.

10.1002/chem.202100252

WILEY-VCH

FULL PAPER

Prediction and Validation of a Druggable Site on Virulence Factor of Drug Resistant *Burkholderia cenocepacia*

Kanhaya Lal,[a][b] Rafael Bermeo,[a][b] Jonathan Cramer,[c] Francesca Vasile,[a] Beat Ernst,[c] Anne Imberty,*[b] Anna Bernardi,*[a] Annabelle Varrot,*[b] Laura Belvisi*[a]

[a] K. Lal, R. Bermeo, Prof. Dr. F. Vasile, Prof. Dr. A. Bernardi, Prof. Dr. L. Belvisi
Universita' degli Studi di Milano, Dipartimento di Chimica
via Golgi 19, I-20133, Milano, Italy
E-mail: anna.bernardi@unimi.it, laura.belvisi@unimi.it

[b] K. Lal, R. Bermeo, Dr. A. Imberty, Dr. A. Varrot,
Université Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France
E-mail: Anne.imberty@cermav.cnrs.fr, annabelle.varrot@cermav.cnrs.fr, Twitter : @AnneImberty

[c] Dr. J. Cramer, Prof. Dr. B. Ernst
University of Basel, Department of Pharmaceutical Sciences
Klingelbergstrasse 50, 4056 Basel, Switzerland

Abstract: *Burkholderia cenocepacia* is an opportunistic gram-negative bacterium that causes infections in patients suffering from chronic granulomatous diseases and cystic fibrosis. It displays significant morbidity and mortality due to extreme resistance to almost all clinically useful antibiotics. The bacterial lectin BC2L-C expressed in *B. cenocepacia* is an interesting drug target involved in bacterial adhesion and subsequent deadly infection to the host. We solved the first high resolution crystal structure of the *apo* form of the lectin N-terminal domain (BC2L-C-nt) and compared it with the ones complexed with carbohydrate ligands. Virtual screening of a small fragment library identified potential hits predicted to bind in the vicinity of the fucose binding site. A series of biophysical techniques and a X-ray crystallographic screening were employed to validate the interaction of the hits with the protein domain. The X-ray structure of BC2L-C-nt complexed with one of the identified active fragments confirmed the ability of the site computationally identified to host drug-like fragments. The fragment affinity could be determined by titration microcalorimetry. These structure-based strategies further provide an opportunity to elaborate the fragments into high affinity anti-adhesive glycomimetics, as therapeutic agents against *B. cenocepacia*.

Introduction

Antimicrobial resistance enables pathogens to resist to the effects of an antibiotic or drug that would usually kill them or limit their growth.[1] The emergence and spread of multidrug-resistant bacteria have challenged the existing treatment regimen which has enormous implications for worldwide healthcare delivery and community health.[1-2] *Burkholderia cenocepacia* is a Gram-negative bacterium belonging to a group of more than 20 species called *Burkholderia cepacia* complex (BCC).[3] BCC species survive in natural sources including water, soil and vegetation. In Nature, BCC bacteria can have both beneficial and detrimental effects on plants[4] but they are also identified as opportunistic human pathogens. In particular, *B. cenocepacia* is responsible for deadly infections in patients with immunocompromised conditions like chronic granulomatous diseases[5] and cystic fibrosis.[6] The treatment of the infection is really challenging, as *B. cenocepacia* strains show extreme resistance to almost all clinically useful antibiotics[7] and cause significant morbidity and mortality. *B. cenocepacia* produces a large number of virulence factors that play an important role in host cell infection.[8] Among them, four soluble lectins (BC2L-A, -B, -C and -D) have been identified, displaying very high sequence similarity with the virulence factor LecB (PA-IIL) from *Pseudomonas aeruginosa*.[9] LecB forms a tetramer with high affinity for fucose,[10] while BC2L-A is a dimer with significant affinity for mannose and oligomannose-type N-glycans.[9, 11] Except BC2L-A, the other three *B. cenocepacia* lectins present additional N-terminal domains.[11-12] For BC2L-C, the C-terminal domain (LecB like) specifically binds to mannose, while the N-terminal domain (BC2L-C-nt) has been structurally characterized as a novel fucose-binding domain with a trimeric TNF-α-like architecture.[13] Thus, BC2L-C represents a novel type of superlectin with dual specificity for fucose and mannose in the N- and C-terminal domains, respectively.[14] BC2L-C as a virulence factor binds to carbohydrates present on the epithelial cells of the host. BC2L-C-nt has higher affinity for fucosylated oligosaccharides and its complexes with H-type 1 and Globo H (H-type 3) oligosaccharides have been recently solved.[15] The super lectin is proposed to be involved in adhesion and inflammation processes.[14] Bacterial adhesion represents the first step of infection, it also enables bacteria to have access to nutrients and to better resist to immune factors, bacteriolytic enzymes and antibiotics.[16] Therefore, preventing glycoconjugate-lectin interactions by anti-adhesive therapy can counteract the infection process at its initial stage.[16-17] This inhibition can be achieved by means of carbohydrate-based synthetic molecules which can compete for the lectin working as antagonists and, thus reduce the level of infection.

Here we describe the development of a structure-based approach to the design of such antagonists. First, we solved the crystal structure of apo BC2L-C-nt and compared the protein surface and bound water molecules with the fucose-bound structure. The X-ray crystal structure in complex with methylseleno-α-L-fuco-pyranoside (MeSe-α-L-Fuc, PDB code 2WQ4) was then used for virtual screening of a small fragment library in the vicinity of the

Accepted Manuscript

This article is protected by copyright. All rights reserved.

10.1002/chem.202100252

fucose-binding site. This procedure identified a region (region 'X') that was most likely to host potential hits. The results were analyzed with the main objective of identifying suitable fragments that docked in region X and could be chemically connected to the fucose core to obtain high-affinity ligands. The interaction of the fragments with the protein domain was confirmed using a group of biophysical techniques including STD-NMR.[18] ITC and X-ray crystallography performed on one fragment confirmed binding at the expected location and therefore the ability of site X to host drug-like fragments. This study provides the rational design tools to elaborate the selected fragments into high-affinity ligands.

Results and Discussion

Analysis of the binding site in crystal structures

Crystal structure of trimeric BC2L-C-nt complexed with, H-type 1 and Globo-H oligosaccharides are available[14-15] revealing three sugar binding sites located at the interface between neighboring chains (A, B, C), and separated by a distance of ~20 Å (Figure 1A). In each fucose binding site (Figure 1B), the key residues Tyr48, Ser82, Thr83, Arg85 from one chain (e.g. chain A) and Tyr58, Thr74, Tyr75, Arg111 from the neighbouring chain (e.g. chain C) play an important role in ligand binding. In addition, two water molecules bridge the sugar and the protein. Both water molecules are conserved in the available X-ray structures of BC2L-C-nt in complex with fucoside and fucosylated oligosaccharides.[14-15] One is deeply buried in the binding site and sandwiched between the protein and the ligand, forming an H-bonding interaction with the HO-3 of fucose (Figure 1B). The second water molecule is more exposed to the solvent and mediates an H-bonding interaction between HO-2 of fucose and the side chain of Tyr58.

The crystal structures of complexes evidenced some promising pockets on the protein surface near the fucose binding site. The occurrence of such pockets in the apo-protein needed to be verified and therefore, we solved the crystal structure of the apo form of BC2L-C-nt at high resolution (1.5 Å). The asymmetric unit in the $P6_3$ space group contains one monomer and crystal symmetry was applied to build the trimer for comparison with other structures. Root-mean-squares values of 0.21 Å and 0.24 Å were obtained when comparing with the trimer complexed with MeSe-α-L-Fuc and Globo-H, respectively. Comparison of binding sites (Figure 1C) did not display significant differences in the amino acid of fucose binding pockets and environment. Only a minor difference is observed at the surface loop (Asn52-Phe54), that is involved in the interaction with methyl group of N-acetylgalactosamine (GalNAc) in the complex with Globo-H (Figure S1). Likewise, small differences in the conformation of the N- and C-terminal residues were noticed due to H-bond interaction between them. The changes in the conformation of the termini further caused a small displacement (0.6 to 1.0 Å) of surface loops (Val28-Asp35, Asp 95-Val100). Analysis of water molecules involved in bridging fucose to protein indicated that the more buried one (W1) is conserved in all structures, while the more exposed one (W2) moves by 1.9 Å in the apo structure.

The new crystal structure therefore confirms that the region surrounding the fucose binding site is of interest for drug design. This surface was analyzed for druggability (ligandability)[19] using the SiteMap[20] tool. SiteMap creates a grid of points based on the depth, size, van der Waals interaction energy, hydrophilicity and hydrophobicity to determine the druggability of a protein region. Based on these characteristics, a single scoring function called SiteScore is assigned to potential druggable regions. For BC2L-C-nt the calculations identified three regions, which we labelled X and Y and Z (Figure 1D) in the vicinity of the fucose. Region Y, consisting of residues Ser82, Thr83 and Phe54 in each monomer, corresponds to the area where larger, fucosylated

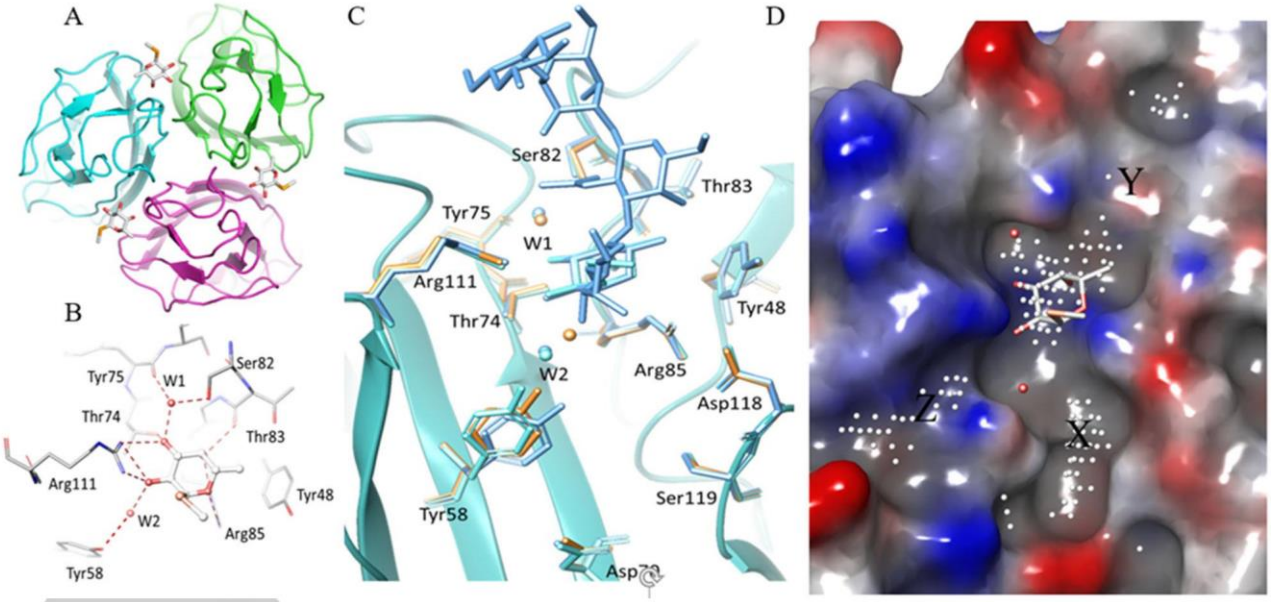


Figure 1 A) Crystal structure of BC2L-C N-terminal domain (PDB 2WQ4) showing three identical fucoside binding sites at the interface of monomers B) Fucoside binding site with MeSe-α-LFuc. Hydrogen bonds are represented as dashed lines. C) Superimposition of binding sites in the apo (orange, PDB 7BFY) and the holo forms (Cyan PDB 2WQ4, azure (PDB 6TIG) of BC2L-C N-terminal domain. D) Identification of additional regions (site points) near fucoside binding site suitable for fragment binding.

Accepted Manuscript

This article is protected by copyright. All rights reserved.

10.1002/chem.202100252

oligosaccharides were observed to bind, including the recently described Globo H hexasaccharide and H-type 1 tetrasaccharide.[15] Of the two other regions (X and Z), site X is a deep crevice extending along the binding interface. The site Z consists of the region between Val110 and Arg111. All the sites are worth exploring further, thus the docking protocol was built to include them in the analysis.

Identification of top-ranked fragments

Docking analysis

2000 molecular fragments were retrieved from the Maybridge library of small fragments (rule of 3 diversity set available at https://www.maybridge.com/). In the first docking model, all the fragments were docked in the presence of the two conserved water molecules and the MeSe-α-L-Fuc. In the second docking model, only the buried water molecule and the ligand were retained since the second water molecule is close to region X and rather exposed to the solvent. This second model allowed us to examine the fragments that might be able to replace it. Fragments were found to dock mainly in regions X and Y. The region Y forms a very shallow and exposed binding site, which mostly hosted lipophilic fragments on the surface. Similarly, region Z also hosted a few hydrophilic fragments on the shallow surface. Region X is comparatively deeper and fragments appear to be nestling in it, generating some specific interactions. Therefore, we focused our further efforts on this region.

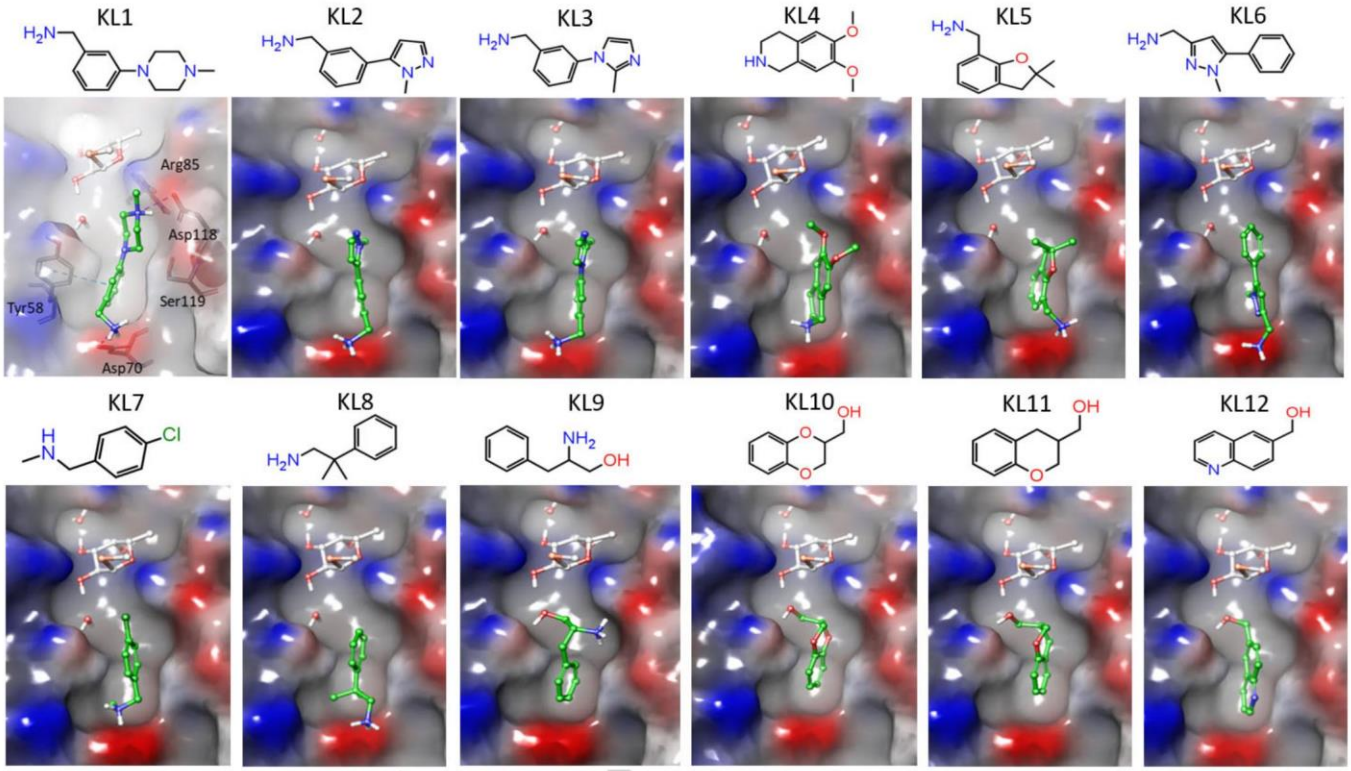


Figure 2 Binding pose for the top ranked fragments (KL1-KL12) predicted by docking studies at site X. The key residues identified in the binding site are shown in the docking pose of KL1.

Binding analysis of top 200 fragments was done for 6 docking runs with XP, SP and HTVS protocols[21] and involving either one water or two water molecules. HTVS and SP use the same scoring function but the HTVS protocol reduces the number of intermediate conformations, torsional refinement and sampling. The XP protocol employs a different, more complex scoring function with greater requirements for ligand-receptor shape complementarity. This screens out false positives that SP or HTVS may let through. From each model, the best fragments with consensus scoring (ranked within top 200 fragments) obtained by XP, SP and HTVS were selected for analysis of key residues involved in ligand binding. The docking results with the two waters model showed that the number of hits obtained at site X using SP and HTVS methods were almost same, while the hits obtained using XP were reduced to half. In the one water model, the number of hits at site X increased almost by a factor of two, due to the omitted water molecule near site X. The interaction pattern identified using three scoring functions at site X indicated that the fragments including a benzylamine moiety have good binding affinity. The key residues involved in binding are Tyr58, and Asp70 whilst Asp118 from neighbouring protomer can also be recognized by some of the top scoring fragments that form a salt bridge interaction with it (Figure 2).

The main interactions observed for the majority of the top ranked fragments are a salt bridge between Asp70 side chain and the benzylamino group of the fragments and π-π stacking interactions with Tyr58. A total of 94 and 89 fragments for site X were identified for one and two waters models, respectively, as top ranked fragments according to XP and SP/HTVS or all the three scoring functions.

This article is protected by copyright. All rights reserved.

Accepted Manuscript

10.1002/chem.202100252

Selection of best fragments

The fragments were carefully analyzed based on different parameters such as structural diversity, possibility to connect them to the fucose core, size and distance from the fucose core. Small fragments which were found significantly far (>6 Å) from the fucose core and docked on the shallow surface surrounding site X were discarded. The remaining 32 fragments for site X were redocked to analyze the stability of the ligand interactions in multiple binding poses (10 poses). Other factors like commercial availability, synthetic feasibility and purchasing cost allowed to select 12 fragments (Figure 2 and Table S1) for experimental validation. Within this group, fragments KL1-8 were among the top scorer in the two waters model, while fragments KL9-12 were predicted to bind in the one water model.

Experimental validation of fragment binding

For each fragment, a 2.5 mM solution was used to test the interaction with BC2L-C-nt using thermal shift assay (TSA, ThermoFluor).[22] Methyl α-L-fucoside (Me-α-L-Fuc) was used as a reference in the experiment to observe fucose binding and hence validate the protocol. Then, the fragments were tested in the presence of Me-α-L-Fuc (20 mM). The results show the expected positive shift (~2 °C) upon Me-α-L-Fuc binding (Figure S2) while all of the complexes with fragments exhibit a small negative shift between 0.15 to 1.65°C (Figure S3) in the melting temperature (T_m), which possibly suggests that the fragments destabilize the binding interface and bind to a non-native or partially unfolded state of the protein.[23]
We repeated the experiment for all the fragments in the absence of Me-α-L-Fuc and the results show similar behavior with a smaller negative shift in the melting temperature (Figure S4). The experimental results of TSA do not afford any structural information concerning the interaction. Therefore, we performed another screening using STD-NMR and X-ray crystallography.

STD-NMR analysis of fragment binding

Saturation transfer difference (STD) NMR has become a leading technique to characterize fragment−macromolecule interaction in solution, because it is sensitive to weak binding events (dissociation constant in a low μM to mM range).[18, 24] In general, STD experiments are performed by irradiating the methyl group of valine, leucine, or isoleucine residues (between 1 and -1 ppm), that are often present in the binding site of proteins.[18]
The irradiation frequency of STD can also be varied in order to investigate whether the fragment has a preferred interaction with aliphatic or aromatic amino acids of the protein.[25]
STD-NMR was used to analyze the interaction of BC2L-C-nt with fragments KL3, KL8 and KL9 in the presence of Me-α-L-Fuc, irradiating at -0.05 ppm. Me-α-L-Fuc was initially tested alone in the experiment, verifying that it binds BC2L-C-nt, with a strong involvement of the methyl group (Figure S5). Then, fragment KL3, KL8 (among the top scorers in the two waters docking model) and fragment KL9 (predicted to bind by the one water model) were analysed in the presence of the protein and of 2 mM Me-α-L-Fuc. The sample was prepared at 1:1 ratio between sugar ligand and fragment. The resulting spectra for fragment KL9 and KL3 are shown in Figure 3 and Figure 4B, respectively. The spectra of fragment KL8 are reported in the supplementary information (Figure S6). In all cases, simultaneous interaction of the fragment and Me-α-L-Fuc with BC2L-C-nt was observed, confirming the binding event for the three fragments in the BC2L-C-nt/fucose complex. In the STD spectra, the signals of Me-α-L-Fuc and of the fragment appear with comparable intensities, indicating a similar affinity for sugar and fragment.

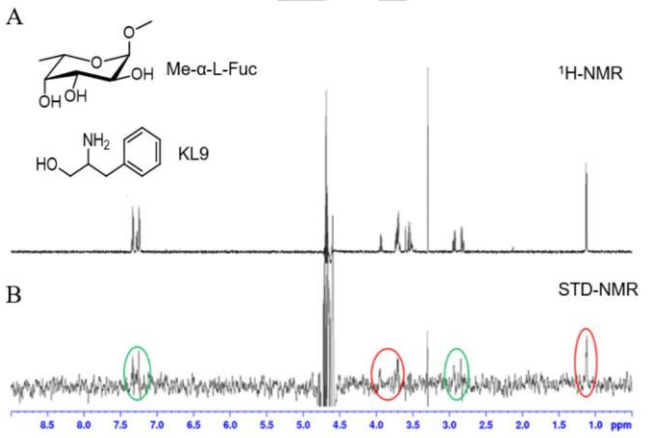


Figure 3. A) 1H-NMR and B) STD spectrum of fragment KL9 and Me-α-L-Fuc in the presence of BC2L-C-nt (1000:1) recorded with a Bruker Avance 600 MHz spectrometer. The spectrum is recorded at 298K with irradiation frequency at -0.05 ppm. In the STD spectrum, the signals at 3.7 ppm and 1.1 ppm, produced respectively by the fucose ring and by its methyl group, are highlighted with red circles. The signals of the fragment are highlighted with a green circle (at 2.9 ppm for -CH2-Ph and 7.3 ppm for the aromatic protons .

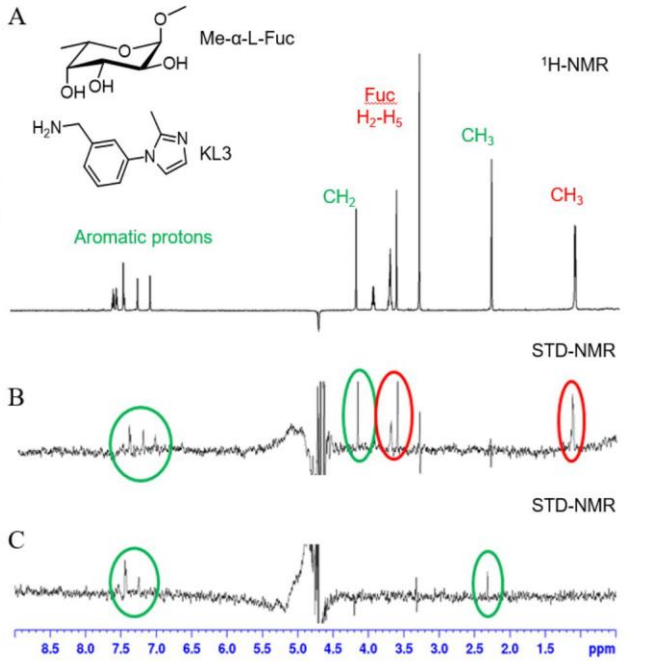


Figure 4 A) ¹H-NMR spectrum, B) STD spectrum (irradiation frequency -0.05 ppm) and C) STD spectrum (irradiation frequency 10 ppm) of fragment KL3 and Me-α-L-Fuc in the presence of BC2L-C-nt (1000:1) recorded with a Bruker Avance 600 MHz spectrometer at 298K. In the STD spectrum at -0.05 ppm (B), the signals at 3.7 ppm and 1.1 ppm, produced respectively by the fucose ring and by its methyl group, are highlighted with red circles. The signals of the fragment are highlighted with a green circle (at 4.2 ppm for -CH_2- and in the range 7.05-7.4 ppm for aromatic protons). The STD spectrum at 10 ppm (C) shows the aromatic protons (in the range 7.1-7.4 ppm) and the methyl group (at 2.3 ppm) of the fragment

Accepted Manuscript

This article is protected by copyright. All rights reserved.

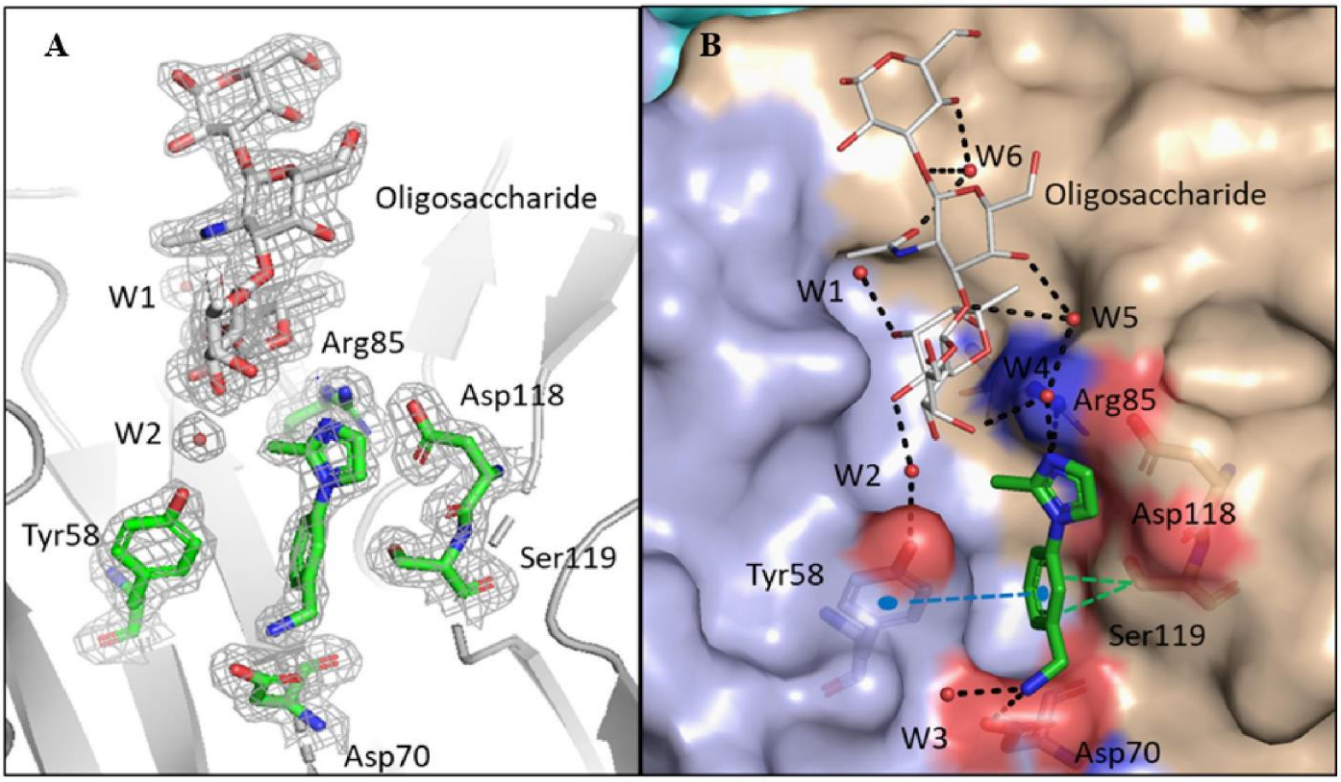


Figure 5. Crystal structure of BC2L-C-nt with Globo H and KL3. A) Zoom in the binding site with 2Fo-DFc electron density represented at 1σ B) Network of interaction in the binding site. Analysis of the complex shows that the key interactions and residues predicted from docking studies were involved in the ligand (KL3) binding. The salt bridge between Asp70 side chain and benzylamino group and π-π stacking interactions with Tyr58 are maintained in the crystallized complex. In addition to the water molecules from two waters model, a new network of water molecules involved in key interactions between the ligand and the protein is also highlighted. H-bonding interactions and hydrophobic interactions are displayed in black and green dashed lines respectively. π-π stacking interactions are shown in blue dashed lines

STD spectra were also acquired using 10 ppm as irradiation frequency. In this case, the aromatic protons of the fragments are observable, while no signals of Me-α-L-Fuc can be detected (Figures 4C and Figure S7). This finding suggests that the fragments bind in the proximity of aromatic residues of the protein and thus supports the docking prediction that they are located in a protein binding pocket that includes an aromatic residue (Tyr58). It is interesting to note that the STD spectrum of KL3 obtained irradiating at 10 ppm (Figure 4C) also shows a clear signal for the methyl group of the fragment (at 2.3 ppm), which is not visible when irradiating at -0.05 ppm (Figure 4B). This suggests that also this moiety is proximal to an aromatic side chain of the protein. On the contrary, the singlet at 4.2 ppm, corresponding to the methyleneamino benzylic protons of the fragment, which is clearly visible when irradiating at -0.05 ppm (Figure 4B), disappears from the spectrum, like the fucose protons, when irradiating at 10 ppm. Thus, this moiety is expected to be surrounded by aliphatic protons of the protein.

KL3-BC2L-C-nt crystal structure analysis

All fragments (KL1-KL12) were soluble enough to be used for soaking experiment with crystals of BC2L-C-nt complexed with Globo H hexasaccharide obtained as described previously.[15] After soaking, crystals containing KL10, KL11 and KL12 did not diffract at sufficient resolution for data collection. Crystal soaked with the remaining fragments (KL1-KL9) diffracted at a resolution close to 2 Å or better, but examination of the electron density after molecular replacement only revealed electron density for the sugar and not for the fragment indicating that they did not bind to the protein in the experimental conditions used. Only in the complex with KL3 (3-(2-Methyl-1H-imidazol-1-yl) benzylamine) at 1.9 Å resolution, electron density corresponding to the expected two monomers. The orientation of the fragment, and the observed interactions correspond very well with those predicted by the docking studies (Figure 5 and Figure S8). Residue Tyr58 forms T-shaped π-π stacking interactions with the benzene ring and Asp70 forms a salt bridge with the amino group in the fragment. The free nitrogen of the imidazole ring makes water mediated interaction with the side chain of Arg85 and the OH-4 of the GlcNAc moiety of Globo H (Figure 5B). The fragment binds with identical pose and reproduce the same binding interactions in the three binding sites of the trimer (Table 1).

The position of the fragment in site X is fully consistent with the STD-NMR data (Figure 4), which indicate proximity of the benzylic methyleneamino group of KL3 to aliphatic residues of the protein (Asp70 and Ser119 in the X-ray structure). The methyl group of KL3, which in the STD spectra responds to irradiation at 10 ppm, is in fact close to the Tyr58 side chain in the X-ray structure. The

Accepted Manuscript

This article is protected by copyright. All rights reserved.

10.1002/chem.202100252

water mediated interactions with Globo H were identical to the previous complex.[15] The results of X-ray crystallographic screening validated the docking results and the ligandability of site X.

Table 1. Summary of the interactions of BC2L-C-nt with KL3 in three binding sites.

Ligand atom	Protein or water atom	Distance (Å)
N3	Asp70 (OD2)	3.20
	W3 (HOH161)[a]	2.75 ± 0.15
N1	Arg85 (NH2)	3.30 ± 0.07
	W4 (HOH108)	2.46 ± 0.05
C[b]	Ser119 (CB)	3.60 ± 0.04
	Tyr58 (CE1)	3.50 ± 0.07

[a] only present in two binding sites

[b] For hydrophobic contacts and π-π interactions, the distance is calculated from the nearest atoms in the ligand and the protein. Mean distance and standard deviation were calculated from the distance of ligand and protein atoms in each binding site.

Affinity analysis and activity validation

The affinity of BC2L-C-nt for KL3 was determined by isothermal titration calorimetry (ITC) measurements.[26] Titration of the lectin by KL3 resulted in small exothermic peaks after correction for buffer mismatch (Figure S9) The integrated curve could be fitted with one-site model with stoichiometry of one, resulting in the determination of a K_d of 877 μM. Because of the low c-value of the experiment, the thermodynamic contributions cannot be safely estimated.

Conclusions

Crystal structure analysis of the apo and the holo form of BC2L-C-nt demonstrated the presence of a druggable (ligandable) region (site X) in the vicinity of the fucoside binding site. The computational and experimental screenings identified fragments interacting with BC2L-C-nt. The study indicates that the fragments bind in a newly identified binding region in BC2L-C-nt when the fucoside binding site is occupied. Different biophysical techniques including TSA and STD-NMR, confirmed fragment-protein interaction. Remarkably, the binding mode of one fragment (KL3) could be validated by X-ray crystallography at high resolution, further confirming the ability of site X to host drug-like fragments. The affinity measured by ITC is sub-millimolar, which is very promising for such small fragment. The complementary structural and thermodynamic data give clear view of the relative importance of apolar and polar interactions for fragment KL3. This could be used in the future for structure-based optimization of this first hit.

Most interestingly, this study provides an opportunity to connect the best fragments to the fucose core to obtain high affinity glycomimetic ligands. The selection of suitable linkers can be done based on the distance (measured 4.8 Å) between the nearest atoms of fragment and the fucose core. Other factors like synthetic feasibility and possibility to maintain the binding pose at the site X can be considered to identify suitable linkers. A robust synthetic route to glycomimetics comprising fucose linked fragments will help in designing high affinity ligands as anti-adhesive agents against *B. cenocepacia*.

Experimental Section

Protein expression and purification

Protein production and purification of the BC2L-C-nt was performed as described previously.[15] An average yield 5.2 mg.L−1of culture medium was obtained and stored at 4 °C.

Preparation of protein model

All the calculations were performed using the Schrödinger Suite through Maestro (version 2018-1) graphical interface.[27] Atomic coordinates from the crystal structure of BC2L-C-nt complexed with MeSe-α-L-Fuc (PDB code 2WQ4) were taken from the Protein Data Bank.[28] The asymmetric unit contains three peptide chains and three carbohydrate ligands (MeSe-α-L-Fuc), around a 3-fold pseudo axis of symmetry. The mode of binding for the sugar is identical in the three binding sites, therefore only one binding site located between chains A and C was used for the calculations. The two structural water molecules HOH2195 (W1) and HOH2194 (W2) bridging fucose and protein were also retained. The hydrogen atoms were added and pKa was predicted for protein residues using the PROPKA[29] method at pH 7.4 and assigned HIE protonation state to the histidine (His116) residue. Finally, protein-ligand complex was subjected to restrained minimization with convergence of heavy atoms to an RMSD of 0.3 Å using the OPLS3 force field.[30]

Preparation of ligand models

The Maybridge library of small fragments (rule of 3 diversity set) containing 2000 fragments was used for *in silico* screening. The LigPrep[31] tool was used to generate tautomers, stereoisomers and protonation states at pH 7±2. The calculation generated 2904 structures.

Models for docking study

For docking grid generation, the centroids of residues from chain A (Tyr48, Ser82, Thr83, Arg85) and chain C (Tyr58, Thr74, Tyr75, Arg111) were selected to define a cubic grid box of 32×32×32 Å. The ligand (MeSe-α-L-Fuc) and the water molecules (HOH2194 and HOH2195) were retained. The same residues were used to generate the second grid with one water (HOH 2195) molecule. Both the grids (models) were used for docking studies using XP, SP and HTVS scoring functions. All the calculations were accomplished by Glide (version 7.8)[21] using the flexible docking approach.

Thermal shift assay (TSA)

The fragments KL1, KL2, KL3, KL5, KL6, KL7, KL9, KL10, KL11 (Table S1) were purchased from the Maybridge (Fisher Scientific International) and the other fragments; KL4, KL8 and KL12 were purchased from the abcr GmbH. The fragments were tested for the purity using liquid chromatography-mass spectrometry (LC-MS).

For the dye-based TSA, BC2L-C-nt (5 μM) in assay buffer (20 mM Tris HCl, 100 mM NaCl, pH 8.0) was incubated with 50x SYPRO orange and 2.5 mM KL1-12 in the presence or absence of 20 mM Me-α-L-Fuc. A Qiagen

Accepted Manuscript

This article is protected by copyright. All rights reserved.

Rotor-Gene Q instrument was used to apply a heat ramp of 1 °C/min from 25-95 °C and SYPRO orange fluorescence at 620 nm was monitored using the appropriate optical channel.

STD-NMR interaction studies

The interaction between ligands and isolated protein was investigated using STD-NMR experiments. The spectra were acquired with a Bruker Avance 600 MHz instrument at 298 K, in a 3 mm NMR tube and in the phosphate buffer previously described (200 µl). All protein−ligand samples were prepared in a 100:1 and 1000:1 ligand/protein ratio in concentration. In STD experiments water suppression was achieved by using the WATERGATE 3-9-19 pulse sequence. The on-resonance irradiation of the protein was kept at -0.05 ppm and 10 ppm. Off-resonance irradiation was applied at 200 ppm, where no protein signals were visible. Selective presaturation of the protein was achieved by a train of Gauss shaped pulses of 49 ms length each. The total length of the saturation train depends on the L7 parameter (the loop counter). STD experiments were acquired with L7 = 60 leading 2.94 s of total saturation. Two protocols for sample preparation were followed in all cases: either by adding the fragment to a pre-incubated solution of protein and Me-α-L-Fuc, or by adding the fucoside to a pre-incubated solution of protein and fragment. The resulting STD spectra were very similar independent of the set up. So, the results reported here correspond to the experiments obtained by adding the fragments to a solution of protein and Me-α-L-Fuc.

X-ray crystallography, data collection, and structure determination

The apo form of BC2L-C-nt was crystallized using the vapour diffusion method and 2 µL hanging drops containing a 50:50 (v/v) mix of protein (5.5 mg/ml) and reservoir (sodium citrate 1.2 M at pH 7.0). Cubic crystals were obtained from the solution after 3 weeks. For the soaking experiments, crystals of BC2L-C-nt in complex with Globo H oligosaccharide were obtained as described previously.[15] The fragments were tested for the aqueous solubility at higher concentration and a stock solution was prepared. The crystals were soaked overnight in the 0.5 µl volume of fragments (from stock) in 4.5µl of 2.5 M sodium malonate used for cryoprotection that makes a final concentration of 2 mM, for the fragments KL1, KL7 and KL11, 2.5 mM for KL12, 5 mM for KL2, KL5, KL6, KL8 and KL10 and 10 mM for KL3, KL4, KL9. For KL2 and KL12, 10 percent DMSO was added to achieve the above concentration. The crystals were flash-cooled in liquid nitrogen prior to data collection. The data was collected on the beamline Proxima 1, synchrotron SOLEIL, Saint Aubin, France, using an Eiger 16 m detector (Dectris, Baden, Switzerland). The data was processed using XDS and XDSME.[32] The CCP4 suite was used for all further processing.[33] The coordinates of the monomer A of PDB code 2WQ4 were used as search model to solve the structures of the apo form and the complexes with BC2L-C-nt by molecular replacement using PHASER.[34] Refinement was performed using restrained maximum likelihood refinement and REFMAC 5.8[35] interspaced with using manual rebuilding in Coot.[36] for cross validation, 5% of the data were set aside. Riding atoms were added during refinement (Table S2). Library for the fragment was made using ligand builder in Coot. All carbohydrates were validated using Privateer in CCP4i2 prior validation using the PDB validation server and deposition to the Protein Data Bank under code 7BFY for the apo form and 6ZZW for the complex..

ITC Measurements

The ITC experiments were performed at 25 °C with an ITC200 isothermal titration calorimeter (Microcal-Malvern Panalytical, Orsay, France). The protein (BC2L-C-nt) and ligand (KL3) were dissolved in the same buffer composed of 100 mM Tris HCl pH 7.0 and 100 mM NaCl. A total of 38 injections of 1 µL of ligand solution (15 mM) were added at intervals of 200 s while stirring at 850 rpm was maintained to ensure proper mixing in the 200 µL sample cell containing the protein, at 225 µM. A control experiment was performed by injecting same concentration of KL3 in buffer. The differences of integrated peaks were performed using the Microcal PEAQ-ITC analysis software. The binding thermodynamics was further processed with a "one set of sites" fitting model. The experiment determined experiment affinity (K_d), binding enthalpy (ΔH) while the stoichiometry was fixed to 1. Free energy change (ΔG) and entropy contributions ($T\Delta S$) were derived from the equation $\Delta G = \Delta H - T\Delta S$. The experiments were performed in duplicates and the standard deviation was in 20% range for K_d.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

This research was funded from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 765581. The authors acknowledge support by the ANR PIA Glyco@Alps (ANR-15-IDEX-02) and Labex Arcane-CBH-EUR-GS (ANR-17-EURE-0003). The authors are grateful to SOLEIL Synchrotron, Saint Aubin, France for provision of synchrotron radiation facilities and access to the beamline Proxima 1. The STD-NMR experiments were performed using the Unitech COSPECT platform at the University of Milan.

Keywords: Glycomimetics • BC2L-C • Ligand design • Virtual screening• Antimicrobial resistance

References

[1] *Antimicrobial Resistance: Global Report on Surveillance.* Geneva: WHO **2014**, pp. 1-252
[2] R. Smith and J. Coast, *BMJ* **2013**, *346*, f1493.
[3] E. Mahenthiralingam, A. Baldwin and C. G. Dowson, *J. Appl. Microbiol.* **2008**, *104*, 1539-1551.
[4] E. Mahenthiralingam, T. A. Urban and J. B. Goldberg, *Nat. Rev. Microbiol.* **2005**, *3*, 144-156.
[5] J. A. Winkelstein, M. C. Marino, R. B. Johnston, Jr., J. Boyle, J. Curnutte, J. I. Gallin, H. L. Malech, S. M. Holland, H. Ochs, P. Quie, R. H. Buckley, C. B. Foster, S. J. Chanock and H. Dickler, *Medicine (Baltimore)* **2000**, *79*, 155-169.
[6] S. L. Butler, C. J. Doherty, J. E. Hughes, J. W. Nelson and J. R. Govan, *J. Clin. Microbiol.* **1995**, *33*, 1001-1004.
[7] S. Nzula, P. Vandamme and J. R. Govan, *J. Antimicrob. Chemother.* **2002**, *50*, 265-269.
[8] a) S. A. Loutet and M. A. Valvano, *Infect. Immun.* **2010**, *78*, 4088-4100; b) M. S. Saldias and M. A. Valvano, *Microbiology* **2009**, *155*, 2809-2817.
[9] A. Imberty, M. wimmerova, E. P. Mitchell and N. Gilboa-Garber, *Microbes Infect.* **2004**, *6*, 221-228.
[10] E. P. Mitchell, C. Sabin, L. Snajdrova, M. Pokorna, S. Perret, C. Gautier, C. Hofr, N. Gilboa-Garber, J. Koca, M. Wimmerova and A. Imberty, *Proteins* **2005**, *58*, 735-746.
[11] E. Lameignere, L. Malinovska, M. Slavikova, E. Duchaud, E. P. Mitchell, A. Varrot, O. Sedo, A. Imberty and M. Wimmerova, *Biochem. J.* **2008**, *411*, 307-318.
[12] E. Lameignere, T. C. Shiao, R. Roy, M. Wimmerova, F. Dubreuil, A. Varrot and A. Imberty, *Glycobiology* **2010**, *20*, 87-98.
[13] O. Sulak, G. Cioci, M. Delia, M. Lahmann, A. Varrot, A. Imberty and M. Wimmerova, *Structure* **2010**, *18*, 59-72.
[14] O. Sulak, G. Cioci, E. Lameignere, V. Balloy, A. Round, I. Gutsche, L. Malinovska, M. Chignard, P. Kosma, D. F. Aubert, C. L. Marolda, M. A. Valvano, M. Wimmerova and A. Imberty, *PLoS Pathog.* **2011**, *7*, e1002238.
[15] R. Bermeo, A. Bernardi and A. Varrot, *Molecules* **2020**, *25*.
[16] I. Ofek, D. L. Hasty and N. Sharon, *FEMS Immunol. Med. Microbiol.* **2003**, *38*, 181-191.
[17] a) B. Ernst and J. L. Magnani, *Nat. Rev. Drug Discov.* **2009**, *8*, 661-677; b) S. Sattin and A. Bernardi, *Trends Biotechnol.* **2016**, *34*, 483-495.
[18] B. Meyer and T. Peters, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 864-890.
[19] S. Vukovic and D. J. Huggins, *Drug Discovery Today* **2018**, *23*, 1258-1266.

Accepted Manuscript

This article is protected by copyright. All rights reserved.

10.1002/chem.202100252

[20] T. A. Halgren, *J. Chem. Inf. Model.* **2009**, *49*, 377-389.
[21] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739-1749.
[22] M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane and F. R. Salemme, *J. Biomol. Screen.* **2001**, *6*, 429-440.
[23] P. Cimmperman, L. Baranauskiene, S. Jachimoviciute, J. Jachno, J. Torresan, V. Michailoviene, J. Matuliene, J. Sereikaite, V. Bumelis and D. Matulis, *Biophys. J.* **2008**, *95*, 3222-3231.
[24] a) F. Vasile, F. Gubinelli, M. Panigada, E. Soprana, A. Siccardi and D. Potenza, *Glycobiology* **2018**, *28*, 42-49; b) F. Vasile, D. Rossi, S. Collina and D. Potenza, *Eur. J. Org. Chem.* **2014**, *2*, 5.
[25] S. Monaco, L. E. Tailford, N. Juge and J. Angulo, *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 15289-15293.
[26] M. R. Duff, Jr., J. Grubbs and E. E. Howell, *J Vis Exp* **2011**.
[27] S. Schrödinger Release 2018-1: Maestro, LLC, New York, NY, **2018**.
[28] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235-242.
[29] a) M. H. Olsson, C. R. Sondergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525-537; b) H. Li, A. D. Robertson and J. H. Jensen, *Proteins* **2005**, *61*, 704-721; c) D. C. Bas, D. M. Rogers and J. H. Jensen, *Proteins* **2008**, *73*, 765-783.
[30] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel and R. A. Friesner, *J. Chem. Theory Comput.* **2016**, *12*, 281-296.
[31] S. Schrödinger Release 2018-1: LigPrep, LLC, New York, NY in *Vol.* **2018**.
[32] a) W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125-132; b) P. Legrand, *GitHub Repos* **2017**, *2017*.
[33] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235-242.
[34] A. J. McCoy, *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 32-41.
[35] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355-367.
[36] P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486-501.

WILEY-VCH
Accepted Manuscript

This article is protected by copyright. All rights reserved.

8.3. Experimental section

The protocols and materials employed for experiments are detailed in this section, they are separated as follows:

1. Production and purification of BC2L-C-N_{ter}
2. Biophysical and structural evaluation of lectins and their interaction
3. Organic synthesis and characterization of small molecules

The principles related to the first two sections are be described in Part 2: **RESEARCH METHODOLOGY**.

1. Production and purification of BC2L-C-N_{ter}

Primer Design, PCR and ligation

The DNA sequence encoding for BC2L-C-N_{ter} comprising amino acids 2 to 132 was amplified by PCR with purposely designed primers using previous construct as template and 5'-CTTCATATGCCGCTGCTGAGCGCCAGTATCG-3' and 5'-TACTCGAGTTATGCCGCGGTGCCCCAAACG-3' as forward and reverse primers, respectively (restriction sites are underlined: NdeI and XhoI respectively). The PCR product (*ca.* 400 base pairs) was purified from 1% agarose gel using Nucleospin Gel and PCR Clean-up kit (Macherey-Nagel, Hoerdt, France) using manufacturer instructions. The gene product and homelab vectors of interest (pET-TEV[207] and pCold-TEV[153]) were digested with NcoI and XhoI restriction enzymes (New England Biolabs, Evry, France) for 1 h at 37 °C, purified, and ligated at room temperature using the Takara mix. The pCold-TEV originates from the pCold-TF vector (Takara Bio Europe, Saint Germain en Laye, France) where the enterokinase site was replaced by tobacco etch virus (TEV) cleavage site by PCR using the 5'-CGCGGTAGTGGTGGTGAAAACCTGTATTTTCAGGGCCATATGGAGCTCGGTACC-3' and 5'-ACCACCACTACCGCGTGGCACCAGACCCGC-3' as forward and reverse primers, respectively, with the PrimeSTAR Max DNA polymerase (Takara Bio Europe, Saint Germain en Laye, France) according to manufacturer instructions.

Protein expression and purification

The aforementioned vector was transformed by heat shock into *Escherichia coli* BL21 Star (DE3) cells. Cells harbouring the plasmid were cultured in Luria Broth (LB) broth medium

supplemented with 100 µg/mL ampicillin at 37 °C under constant shaking at 170 rpm. At OD_{600nm} = 0.4, the incubator temperature was decreased to 16 °C and when OD_{600nm} reached 0.7, the protein expression was induced overnight by the addition of 0.1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Then, cells were centrifuged at room temperature, 5 min at 5000 x *g* and the resulting pellet was weighed and used or stored at -20 °C. Each gram of wet cell pellet was resuspended with 5 mL of Buffer 1 (Tris-HCl 50 mM, NaCl 100 mM, pH 8.5) prior to treatment with DENARASE® endonuclease (c-LEcta GMBH, Leipzig, Germany) for 10 min at room temperature on a rotating wheel. The cells were lysed by pressure at 1.9 MPa using a one-shot table-top cell disruptor (Constant Systems Ltd., UK). The lysate was centrifuged 30 min, 24,000× *g* at 4 °C, and the resulting supernatant filtered through a 0.45 µm polyethersulfone (PES) syringe filter prior to loading on a 5 mL HisTrap™ fast flow (FF) column (GE Heathcare Life Sciences, Marlborough, MA, USA) equilibrated with buffer 1 for affinity chromatography using NGC system (Bio-Rad, Marnes-la-Coquette, France). After washing the unbound proteins with buffer 1, rBC2L-CN2 was eluted using a 20 column volumes (CV) gradient of 0–500 mM imidazole. Fractions containing the protein were pooled after examination on 15% SDS-PAGE gel. The imidazole was removed using a PD10 desalting column (GE Healthcare Life Sciences, Marlborough, MA, USA). The protein was concentrated by centrifugation (Vivaspin 3kDa, Sartorius, Goettingen, Germany) to at least 0.7 mg/mL before addition of TEV protease (1:50 *w*/*w*, enzyme:protein ratio), 1 mM ethylenediaminetetraacetic acid (EDTA) and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) for tag cleavage overnight at 19 °C.[154] The sample was again submitted to affinity chromatography (same conditions as previously) to separate two fragments of 14 kDa and 52 kDa corresponding to the target protein and its cleaved fusion, respectively (assessed by SDS–PAGE 15 %). After concentration by centrifugation as previously described, the protein concentration was determined by UV absorbance at 280 nm with a NanoDrop 2000 spectrophotometer for an extinction coefficient ε = 19940 (Thermo Scientific, Illkirch-Graffenstaden, France).

SEC was performed on an ENrichTM SEC 70 10 × 300 column (Bio-Rad, Marnes-la-Coquette, France) using a NGC™ systems (Bio-Rad Ltd.). The analytical column was pre-equilibrated with 20 mM Tris-HCl pH 7.0 and 100 mM NaCl, optimized for protein stability via thermal shift assay (TSA). The volume for the sample injections was 240 µL and the flow rate was 1.0

mL/min. A column calibration curve using gel-filtration standards (GE Healthcare, Life Sciences) was performed to allow the calculation of the protein molecular weight.

2. Biophysical and structural evaluation of lectins and their interactions

Isothermal Titration Calorimetry

All experiments were performed at 25 °C with an ITC200 isothermal titration calorimeter (Microcal-Malvern Panalytical, Orsay, France). The protein rBC2L-CN2 and its ligands were dissolved in a buffer composed of 20 mM Tris HCl pH 7.0 and 100 mM NaCl. The 200 µL sample cell containing rBC2L-CN (concentrations ranging from 200 to 400 µM) was subjected to injections of ligand solution: 20 to 39 injections of 1 µL or 70 injections of 0.5 µL (5 to 50 mM, chosen depending on the ligand) at intervals of 100, 120 or 200s while stirring at 850 rpm. Control experiments were performed by repeating the same protocol, but injecting the ligand into buffer solution. The supplied software Origin 7 or MicroCal PEAQ-ITC was used to fit the experimental data to a theoretical titration curve allowing the determination of affinity (i.e., association constant, K_a), binding enthalpy (ΔH), and stoichiometry (n). Values for free energy change (ΔG) and entropy contributions (TΔS) were derived from the equation $\Delta G = \Delta H - T\Delta S = - RT \ln K_a$ (with T = 298.15 K and R = 8.314 $J.mol^{-1}.K^{-1}$. For experiments in ligand excess, the stoichiometry was fixed to 1. Two experiments were performed for L-galactose, αMe-L-fucose and the synthetic ligands, three for H-type 1 trisaccharide and Globo H hexasaccharide (H-type 3), and only one for H-disaccharide (Fucα1-2Gal), H-type 1 tetrasaccharide, or Lewis Y pentasaccharide. The oligosaccharides were purchased from Elicityl, Crolles, France.

Surface plasmon Resonance

Experiments were performed on a BIACORE X100 instrument (GE Healthcare) at 25 °C in running buffer 10 mM HEPES pH 7.4, 150 mM NaCl and 0.05% Tween 20, adjusted to include 8% DMSO when indicated. rBC2L-CN2 was immobilized onto CM5 chips (BIACORE) following the amine coupling procedure:

- Activation of the chip by three injections of a NHS/EDC mixture at 10 µL/min for 540s, until a minimum of 300 RU was observed on both channels.

- Injection of rBC2L-CN2 (0.5 mg/mL) dissolved in 10 mM sodium acetate pH 4.5 onto channel 2 (contact time of 540s, flow rate of 10 µL/min), until a minimum of 7000 RU was observed for rBC2L-CN2.
- Inactivation of both channels by injecting a 1M ethanolamine (pH 8.5) solution at 5 µL/min for 1080s, achieving over 400 and 7000 RU for channel 1 and 2, respectively.

The analytes were dissolved in the running buffer at increasing concentrations (range: 3.28 – 3500 µM) and subjected to multi-cycle affinity studies (300s association, 300s dissociation, flow rate 5 µL/min). Injections of compounds at increasing concentrations onto the immobilized rBC2L-CN2 were followed by regeneration of the surface: 10 mM fucose in running buffer, then running buffer at 5 µL/min (100s and 150s, respectively) after each analyte association/dissociation. For the higher concentrations, regeneration was secured by performing one or more runs replacing analyte by running buffer. Duplicates were performed for all ligands except FucAmd13 (8% DMSO) and H-type 1 trisaccharide. Binding affinity (K_d) was measured after subtracting the channel 1 reference (no immobilized protein) and subtracting of a blank injection (running buffer - zero analyte concentration). Data evaluation and curve fitting was performed using the provided BIACORE X100 evaluation software (version 2.0).

Differential Scanning Calorimetry

Experiments were performed on a Microcal PEAQ-DSC instrument (Malvern Panalytical, Orsay, France). A buffer composed of 20 mM Tris HCl pH 7.0 and 100 mM NaCl was used to dilute the protein rBC2L-CN2 and its ligands to concentrations 14.3 µM and 143 µM, respectively. Samples of 250µL were loaded, while the reference cell was filled with the matching buffer (aforementioned buffer, ligands when relevant). Each sample was subjected to a gradient of temperature from 20 to 130 °C, at a scan rate of 200 °C/hr, followed by a second similar gradient, generating a reference thermogram. The data was acquired on 'Low' feedback mode. The supplied software MicroCal PEAQ-DSC Software 1.53 was used to fit the experimental data. To obtain the final thermograms, each experiment had its reference thermogram subtracted, the 'Progress' baseline fitting method was used. The profile obtained

was fitted with a 'NonTwoState' model, accounting for two thermal events. Each experiment was performed in duplicates and their averages were calculated by the software.

Saturation Transfer Difference - NMR

^{1}H STD-NMR spectra were acquired at 283 K on a Bruker AVANCE 600 MHz spectrometer. The protein and ligand were dissolved in phosphate buffer (Na_2HPO_4, KH_2PO_4) 20mM pH 7.4, 100 mM NaCl and 5% D_2O in a 3 mm NMR tube (160 µl). Ligand/protein ratios were adjusted to 1000:1 in molar concentration. Water suppression was achieved by using the WATERGATE 3-9-19 pulse sequence. The on-resonance irradiation of the protein was kept at -0.05 ppm and 10 ppm. Off-resonance irradiation was applied at 200 ppm, where no protein signals were visible. Selective pre-saturation of the protein was achieved by a train of Gauss shaped pulses of 49 ms length each. The experiments were acquired with a saturation time of 2.94 s.

Crystallization, Data Collection, and Structure Determination

Ligand stock solutions (10-50 mM) in water or buffer 20 mM Tris HCl pH 7.0 and 100 mM NaCl were added to rBC2L-CN2 at a concentration of 5 mg/mL, such that the final ligand concentration was 1 mM. After incubation at room temperature (22 °C) for at least 30min, crystallization conditions were screened using the vapor diffusion method and 2 µL hanging drops containing a 50:50 (*v/v*) mix of protein and reservoir solution at 19 °C. The screens used included: BCS Eco Screen, Eco Structure Screen 2, Morpheus I-carboxylic acids, and MIDAS (Molecular Dimensions Ltd., Sheffield, UK). Crystals were obtained in a few days from solution 48 of the Structure Screen 2 and optimized using 1.2–1.4 M tri sodium citrate pH 7.0. Apo crystals were obtained as thick hexagons and complexes led to clusters of plates which were broken to single plates, transferred in 2.5 M sodium malonate (CryoProtX, Molecular Dimensions Ltd.) for cryo-protection, and flash-cooled in liquid nitrogen prior to data collection. Soaking experiments were performed by either soaking an apo crystal in a 2.5 M sodium malonate solution containing ligand at 1-5 mM for 1min prior to flash-cooling, or by transfer of an existing rBC2L-CN2/H-type 1 complex crystal into standard co-crystallization conditions (as described above, ligand at 1.25 M) and incubation at 19 °C for 3 days before flash-cooling.

Data were collected at:

- European Synchrotron Radiation Facility (ESRF), Grenoble, France, on beamline FIP-BM30A using a ADSC Q315r detector (Area Detector Systems Corporation, Poway, CA, USA)
- Synchrotron SOLEIL, Saint Aubin, France on beamline Proxima 1, using an Eiger-X 16M detector (Dectris, Baden, Switzerland) and beamline Proxima 2A (PX2-A), using an Eiger-X 9M detector.

The data were processed using XDS and XDSME and then programs of the CCP4 suite were used.[155-156, 208] The coordinates of PDB-ID 2WQ4 were used as a search model to solve all new structures of rBC2L-CN2 by molecular replacement using PHASER.[209] Refinement was performed by multiple iterations of restrained maximum likelihood refinement and REFMAC 5.8 and manual rebuilding in Coot.[210-211] 5% of the observations was set aside for cross-validation analysis. Hydrogen atoms were added in their riding positions during refinement. The final model was validated at the wwPDB Validation server, https://validate-rcsb-1.wwpdb.org/ and the carbohydrate conformations were checked in Privateer.[212] A library for the synthetic molecules was created in the Coot ligand builder. The coordinates were deposited in the Protein Data Bank (PDB) under codes 6TID and 6TIG for H-type 1 and Globo H complex structures, respectively.

3. Organic synthesis and characterization of small molecules

General

Chemicals were purchased from commercial sources and used without further purification, unless otherwise indicated. When anhydrous conditions were required, the reactions were performed under nitrogen or argon atmosphere. Anhydrous solvents were purchased from Sigma-Aldrich® with a content of water ≤0.005%. Triethylamine (Et_3N), methanol and dichloromethane were dried over calcium hydride, THF was dried over sodium/benzophenone and freshly distilled. N,N-Dimethylformamide (DMF) was dried over 4 Å molecular sieves. Washing solutions used in the work-up procedures were saturated, unless stated otherwise. Reactions were monitored by analytical thin-layer chromatography (TLC) performed on Silica Gel 60 F_{254} plates (Merck), and TLC Silica gel 60 RP-18 F_{254}s (Merck), analysed with UV detection (254 nm and 365 nm) and/or staining with:

- Ammonium molybdate acid solution (ammonium molybdate tetrahydrate: 21 g, cerium (IV) sulfate: 1 g, distilled water: 470 mL, sulfuric acid: 30 mL)
- Potassium permanganate alkaline solution (potassium permanganate: 3 g, potassium carbonate: 20 g, sodium hydroxide 1.25 M: 5 mL, distilled water: 300 mL)
- Ninhydrin stain (ninhydrin: 300 mg, acetic acid: 3 mL, n-butanol: 100 mL)

Flash column chromatography was performed using silica gel 60 (40–63 µm, Merck). Automated flash chromatography was performed with Biotage Isolera Prime system and SNAP ULTRA cartridges were employed. For HPLC purifications, a Waters 600 controller coupled to a Waters 2487 Dual Absorbance Detector (214 and 250 nm) were used at a flow rate of 22.0 mL/min (Varioprep column: 250/21 mm nucleosil 100-7 C_{18}). The gradient used was linear from H_2O (0.1% TFA) to CH_3CN 9/1 H_2O (0.1% TFA). NMR experiments were recorded on a Bruker AVANCE 400 MHz instrument at 298 K. Chemical shifts (δ) are reported in ppm. The ^1H and ^{13}C NMR resonances of compounds were assigned with the assistance of COSY and HSQC experiments. Multiplicities are assigned as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), mult. (for multiplets encompassing more than one proton). Broad peaks are denoted by *b*. Aromatic and heteroaromatic protons and carbons are denoted as *Ar*, or *hAr* when ambiguous.

Mass spectra were recorded on a Thermo Fischer LCQ apparatus (ESI ionization). High resolution mass spectra were recorded on spectrometers Apex II ICR FTMS (ESI ionization-HRMS) or Thermo Fischer LTQ Orbitrap XL (ESI ionization-HRMS). Exact mass analyses were obtained from a VG Autospec M246 (Fisons) spectrometer equipped with EBE geometry and EI source. Specific optical rotation values were measured using either a Perkin-Elmer 241 or a ADP410 polarimeter at 589 nm with a 1.0 or 0.5 dm cell, respectively.

Synthetic protocols and characterizations

Guide:

Synthesis of fucoside building blocks: 183

2 - 8, **17**, **28**, **29**, **56**.

Synthesis and derivatization of fragments (grouped by fragment structure): 206

9, **13**, **21 - 23**, **30**, **34**, **57 - 66**.

Coupling procedures (grouped by type of coupling): 227

10, **14**, **18**, **24**, **25**, **31**, **35**, **38**, **41**, **43**, **46**, **48**, **50**, **52**, **54**.

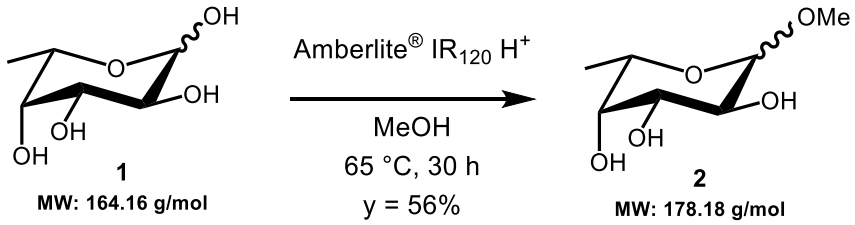
Removal of protecting groups (grouped by protecting group): 260

15, **16**, **19**, **20**, **26**, **27**, **32**, **33**, **36**, **37**, **39**, **40**, **42**, **44**, **45**, **47**, **49**, **51**, **53**, **55**.

Others: 301

68 - 73.

(2) Synthesis and characterization of **Methyl α-L-fucopyranoside (2)** following the procedure of Wang and co-workers:[213]

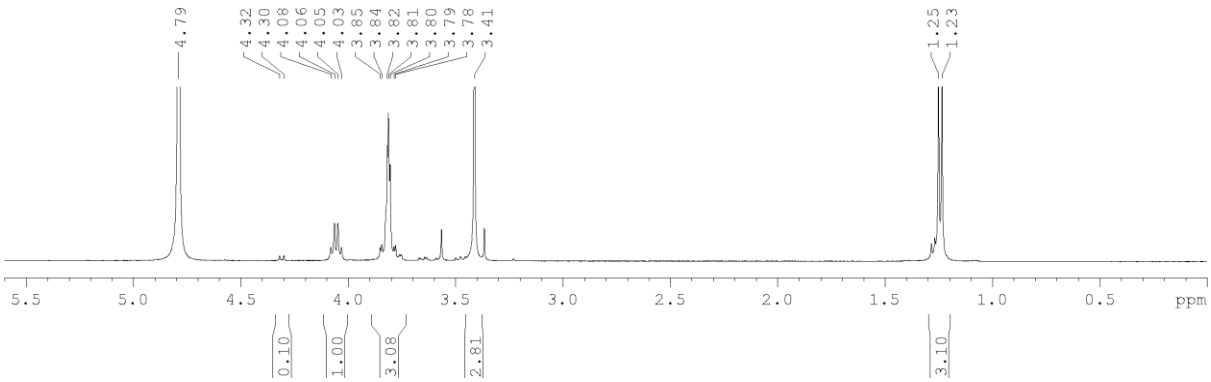


To a solution of L-fucose **1** (5.0 g, 30.46 mmol, 1 eq) in dry MeOH (25 mL) was added Amberlite® IR_{120} H^+ resin (5.0 g) under N_2 atmosphere. The reaction mixture was heated to reflux (65 °C) and stirred for 30 h until TLC showed no further advancement. The reaction was allowed to return to room temperature and was filtered. Evaporation afforded a mixture of Me-α-L-fucopyranoside, Me-β-L-fucopyranoside, Me-α-L-fucofuranoside, and Me-β-L-fucofuranoside (5.2 g), with ratios observed by ^{1}H NMR in agreement with Mowery (53/29/5/10).[182] Recrystallization of Me-α-L-fucopyranoside **2** was achieved by suspending the crude in EtOAc (95 mL: 55 mg/mL), dissolving at reflux (77 °C) and cooling to 5 °C, then filtering white crystals (3.04 g, 17.06 mmol, y = 56%). The evaporation of mother liquors afforded a solid (2.39 g) which was recycled by re-equilibrating in refluxing MeOH (65 °C) with Amberlite® IR_{120} H^+ resin (2.4 g) for 48 h. Further recrystallization from EtOAc and recycling increased the final yield to 80%. TLC R_f (EtOAc/iPrOH/H_2O: 65/35/2.5): 0.52 (Me-α-L-fucopyranoside), 0.65 - 0-75 (Me-β-L-fucopyranoside, Me-α,β-L-fucofuranoside).

^{1}H NMR (400 MHz, D_2O):

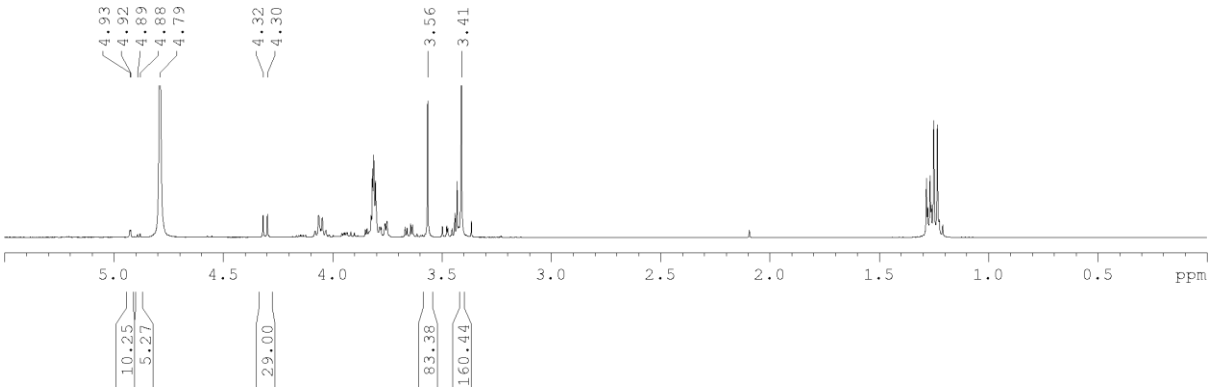
- Crystals of Me-α-L-fucopyranoside **2** (10 % of Me-β-L-fucopyranoside):

δ = 4.77 (1H, *H-1*, masked by solvent), 4.06 (q, $J_{5\text{-}CH3}$ = 6.6 Hz, 1H, *H-5*), 3.85 - 3.78 (mult., 3H, *H-2* + *H-3* + *H-4*), 3.41 (s, 3H, OCH_3), 1.24 (d, $J_{CH3\text{-}5}$ = 6.6 Hz, 3H, CH_3). 4.31 (β fucopyranoside *H-1*): ratio 0.1:1.0. in accordance with published data.[182, 214-215]

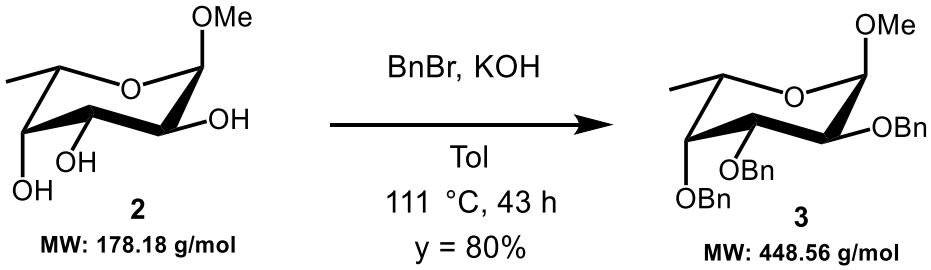


- Crude mixture of 4 isomers:

δ = 4.93 (**β furanoside**, d, *H-1*, J_{1-2} = 1.3 Hz), 4.89 (**α furanoside**, d, *H-1*, J_{1-2} = 4.6 Hz), 4.77 (**α pyranoside**, *H-1*, masked by solvent), 4.31 (**β pyranoside**, d, *H-1*, J_{1-2} = 7.9 Hz), 3.56 (**β pyranoside**, s, OCH_3), 3.41 (**α pyranoside**, s, OCH_3). Ratio/percentage: 10/5/53/29, calculated from H-1 or OCH_3, in accordance with published data.[182, 214-215]



(3) Synthesis and characterization of **Methyl 2,3,4-tri-O-benzyl L-fucopyranoside (3):**

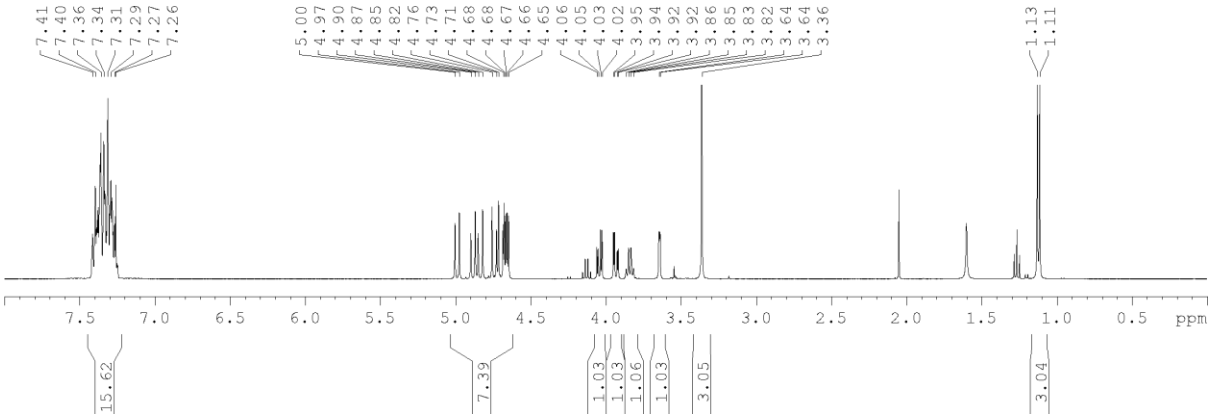


To a solution of **Methyl α-L-fucopyranoside 2** (1.0 g, 5.61 mmol, 1 eq) in dry Toluene (30 mL) was added BnBr (8.0 mL, 67.26 mmol, 12 eq), followed by ground KOH (3.15 g, 5.140 mmol, 10 eq). The reaction mixture was heated to reflux (111 °C) under N_2 atmosphere and stirred for 43 h, before returning to room temperature. The mixture was quenched with ice-cold water and extracted with EtOAc. The organic phase was washed with ice-cold water and dried over $MgSO_4$. The crude was purified by automatic chromatography (Biotage SNAP 100: nHex/EtOAc 5 % isocratic, then gradient to 50 %) affording product **3** (2.02 g, 4.51 mmol, y = 80 %) as an oil. TLC R_f (nHex/EtOAc: 7/3): 0.64. MS (ESI) calculated for $C_{28}H_{32}O_5$ $[M + Na]^+$ *m/z*: 471.21; found: 471.41.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.41 - 7.26 (mult., 15H, *Ar*), 5.00-4.65 (mult., 6H, *CH_2-OBn*), 4.66 (d, J_{1-2}= 3.6 Hz, 1H, *H-1*), 4.04 (dd, J_{2-1} = 3.6 Hz, J_{2-3} = 10.1 Hz, 1H, *H-2*), 3.93 (dd, J_{3-2} = 10.1 Hz, J_{3-4} = 2.8 Hz, 1H, *H-3*),

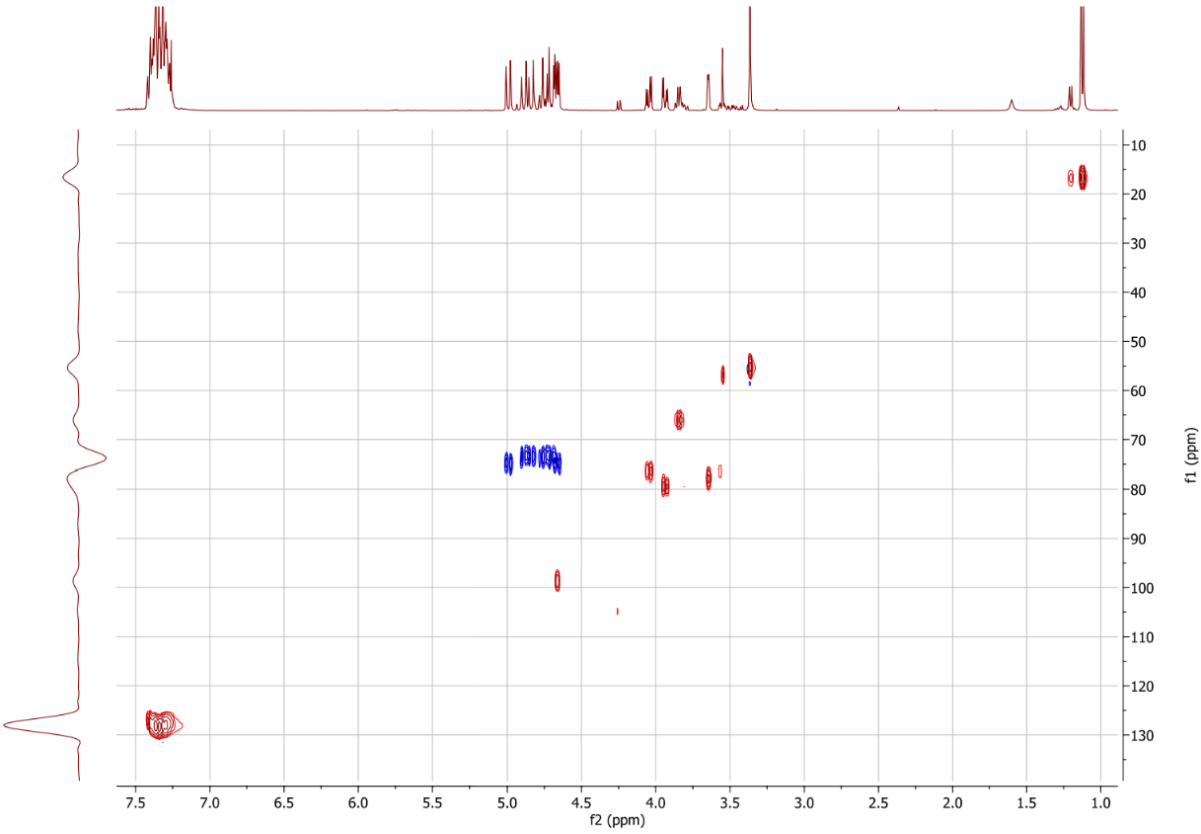
3.84 (dq, $J_{5\text{-}4}$ = 1.3 Hz, $J_{5\text{-}CH3}$ = 6.5, 1H, *H-5*), 3.64 (dd, $J_{4\text{-}3}$ = 2.9 Hz, $J_{4\text{-}5}$ = 1.3 Hz, 1H, *H-4*), 3.36 (s, 3H, *OCH_3*), 1.12 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*). In accordance with published data.[216]



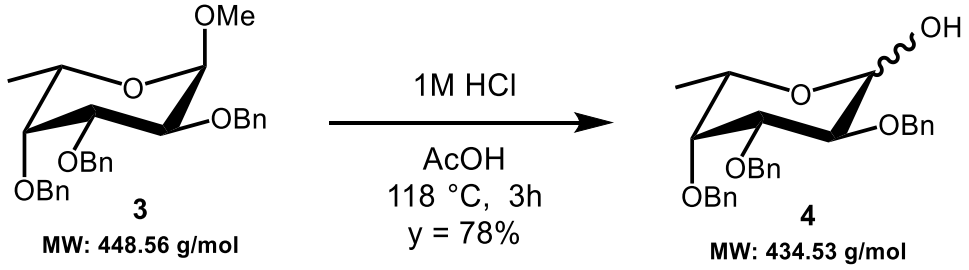
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 128.3-127.4 (*CH Ar*), 98.7 (*C-1*), 79.5 (*C-3*), 77.9 *C-4*), 76.4 (*C-2*), 74.8-73.4 (*CH_2-OBn*), 66.1 (*C-5*), 55.3 (*OCH_3*), 16.6 (*CH_3*).

HSQC (from a reaction using the anomeric mixture, minor product is **Methyl 2,3,4-tri-O-benzyl β-L-fucopyranoside**):



(4) Synthesis and characterization of **2,3,4-tri-O-benzyl L-fucopyranose (4)** following the procedure of Nishi and co-workers:[216]

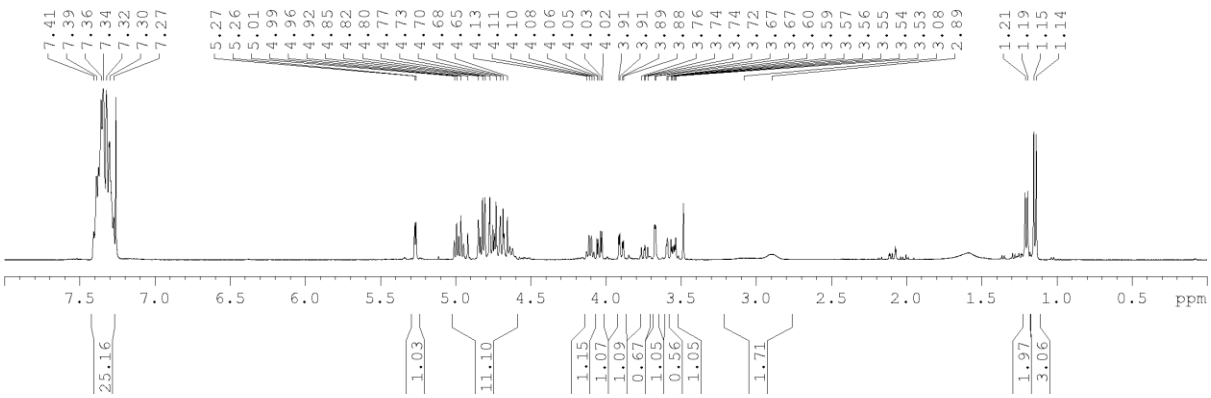


To a solution of **Methyl 2,3,4-tri-O-benzyl L-fucopyranoside 3** (5.6 g, 12.48 mmol, 1 eq) in Acetic acid (100 mL) was added 1M HCl (25 mL, 25 mmol, 2.0 eq). The reaction mixture was heated to reflux (118 °C) and stirred for 3 h until TLC showed completion, before returning to room temperature. The mixture was quenched with ice-cold water and extracted with DCM. The organic phase was washed with $NaHCO_3$ aqueous solution and brine, then dried over Na_2SO_4. Recrystallization was achieved by suspending the crude in a Hex/Et_2O solution (40/13 mL: *ca.* 110 mg/mL), dissolving by heating to 45 °C, cooling to -16 °C, then filtering white crystals of the anomeric mixture **4** (4.24 g, 9.76 mmol, y = 78 %). TLC R_f (nHex/EtOAc: 7/3): 0.34.

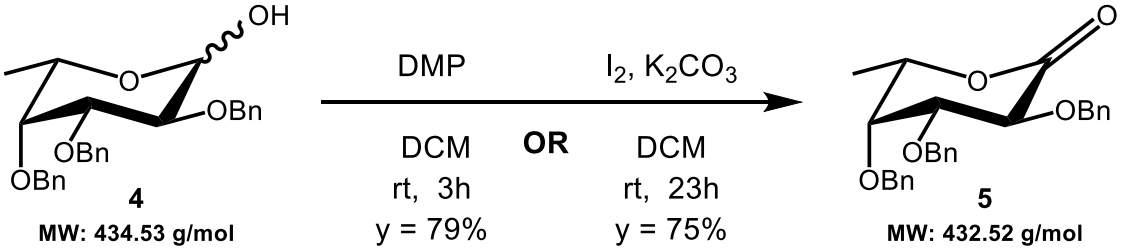
^{1}H NMR (400 MHz, $CDCl_3$): anomeric mixture α/β (ratio 2:1)

α-anomer: δ = 7.36 - 7.27 (mult., 15H, *Ar*), 5.26 (d, J_{1-2}= 3.6 Hz, 1H, *H-1*), 5.01-4.65 (mult., 6H, *CH_2-OBn*), 4.10 (q, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 4.04 (dd, J_{2-1}=3.7 Hz, J_{2-3}=9.9 Hz, 1H, *H-2*), 3.90 (dd, J_{3-2} = 9.9 Hz, J_{3-4} = 2.8 Hz, 1H, *H-3*), 3.67 (dd, J_{4-3} = 2.8 Hz, J_{4-5} = 1.1 Hz, 1H, *H-4*), 2.89 (bs, 1H, *OH*), 1.14 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*). In accordance with published data.[216-217]

β-anomer: δ = 7.36 - 7.27 (mult., 15H, *Ar*), 5.01-4.65 (mult., 6H, *CH_2-OBn*), 4.63 (d, 1H, *H-1*), 3.74 (t, 1H, *H-2*), 3.59 (d, 1H, *H-4*), 3.55 (mult., 2H, *H-3* + *H-5*), 3.08 (bs, 1H, *OH*), 1.20 (d, 3H, *CH_3*).



(5) Synthesis and characterization of **2,3,4-tri-O-benzyl L-fuconolactone (5)**:



Oxidation with Dess-Martin Periodinane (DMP), following the procedure of Frédéric and co-workers:[218]

To a solution of **2,3,4-tri-O-benzyl L-fucopyranoside 4** (242 mg, 0.56 mmol, 1 eq) in DCM (1.6 mL) was added a 15% (weight/weight) DMP in DCM commercial solution (4.0 mL - 817 mg, 1.93 mmol, 3.5 eq) under N_2 atmosphere. The reaction mixture was stirred at room temperature for 3 h until TLC showed completion, before being diluted with DCM and washed with ice-cold $NaHCO_3$ aqueous solution. The organic phase was separated and dried over $MgSO_4$. The crude was purified by automatic chromatography (Biotage SNAP 10: nHex/EtOAc gradient from 5 % to 60 %) affording product **5** (190 mg, 0.44 mmol, y = 79 %). TLC R_f (nHex/EtOAc: 7/3): 0.64. $[\alpha]_D^{19}$ = -79.6 ($CHCl_3$, c 1).

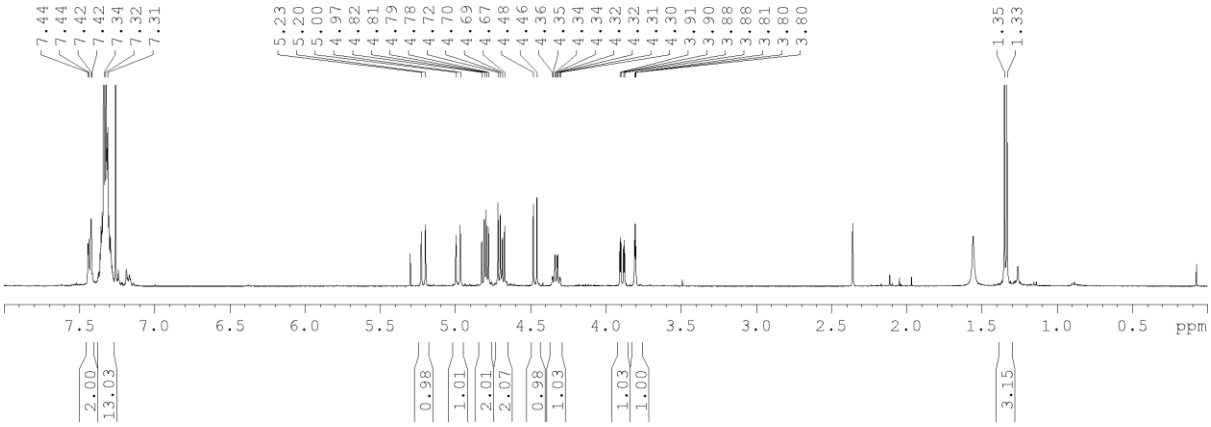
Oxidation with I_2, following the procedure of Fusaro and co-workers:[219]

To a solution of **2,3,4-tri-O-benzyl L-fucopyranoside 4** (4.08 g, 9.39 mmol, 1 eq) in DCM (60 mL) was added I_2 (8.33 g, 32.82 mmol, 3.5 eq), followed by ground K_2CO_3 (4.54 g, 32.85 mmol, 3.5 eq). The brown reaction mixture was stirred at room temperature for 23 h until TLC showed completion, before being quenched with ice-cold water and extracted with DCM. The organic phase was washed with $Na_2S_2O_3$ aqueous solution, becoming clear, and with brine and then dried over Na_2SO_4. The crude was purified by automatic chromatography (Biotage SNAP 100: nHex/EtOAc gradient from 2% to 30%) affording product **5** (3.00 g, 6.94 mmol, y = 74 %). TLC R_f (nHex/EtOAc: 7/3): 0.64.

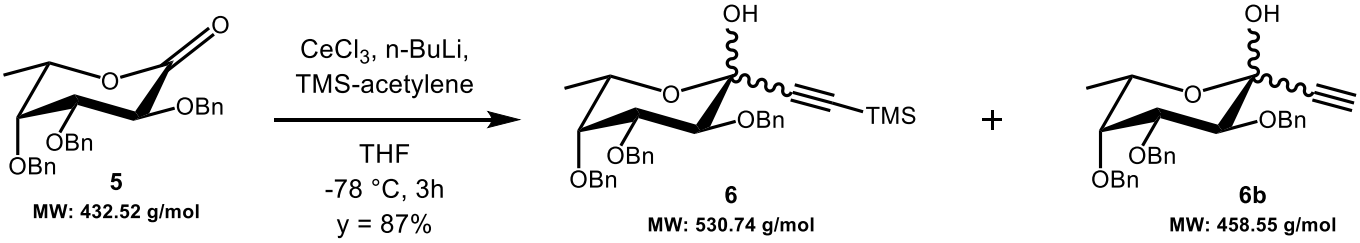
^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.44 - 7.31 (mult., 15H, *Ar*), 5.21 (d, J= 11.0 Hz, 1H, *CH_2-OBn*), 4.98 (d, J'= 11.5 Hz, 1H, *CH_2-OBn*), 4.80 (m, J'''= 11.9 Hz, J'= 11.0 Hz, 2H, *CH_2-OBn*), 4.69 (m, J'''= 12.0 Hz, J'= 11.5 Hz, 2H, *CH_2-OBn*), 4.47 (d, J_{2-3}= 9.6 Hz, 1H, *H-2*), 4.33 (dq, J_{5-4} = 1.5 Hz, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 3.89

(dd, $J_{3\text{-}2}$ = 9.6 Hz, $J_{3\text{-}4}$ = 2.2 Hz, 1H, *H-3*), 3.80 (dd, $J_{4\text{-}3}$ = 2.3 Hz, $J_{4\text{-}5}$ = 1.6 Hz, 1H, *H-4*), 1.34 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*). In accordance with published data.[218]



(6) Synthesis and characterization of **(1-hydroxy-2,3,4-tri-O-benzyl-L-fucopyranosyl) trimethylsilyl acetylene (6)** and **(1-hydroxy-2,3,4-tri-O-benzyl L-fucopyranosyl) acetylene (6b)** following the procedure of Lowary and co-workers:[181]



Oven-dried glassware was used for this procedure. $CeCl_3$ (1.230 g, 4.99 mmol, 2.6 eq) was ground and heated (120 - 140 °C) under high vacuum for 35min. After returning to room temperature, it was flushed with Argon, cooled to 0 °C, and suspended in freshly distilled THF (14 mL). The flask was cooled to -78 °C and left to stir for 2h until the next addition.

Simultaneously, a flask under Ar atmosphere was cooled to -78 °C before adding a 0.9 M solution of TMS-acetylene in dry THF (8.5 mL, 7.65 mmol, 4.0 eq), then a 2.5 M solution of n-BuLi in hexanes (3.4 mL, 8.5 mmol, 4.4 eq). The reaction mixture was stirred at -78 °C for 1h45 before being added to the -78 °C $CeCl_3$ suspension via cannula. The opaque white reaction suspension was stirred at -78 °C for 50 min until the next addition.

Simultaneously, **2,3,4-tri-O-benzyl L-fuconolactone 5** (830 mg, 1.92 mmol, 1 eq) was flushed with Argon, cooled to -78 °C, and dissolved in freshly distilled THF (14 mL). The solution was then added to the -78 °C reaction mixture via cannula, producing a slight peach coloration in the opaque white suspension. The resulting reaction mixture was stirred for 2 h at -78 °C then was allowed to return to room temperature. The reaction was quenched with a 0.1 M HCl aqueous solution and extracted with EtOAc. The organic phase was washed with brine, then

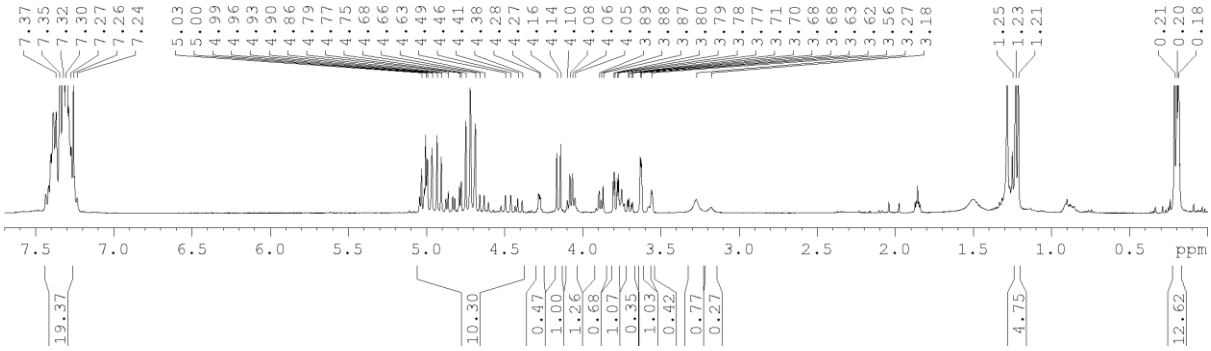
dried over Na_2SO_4. The crude was purified by automatic chromatography (Biotage SNAP 100: nHex/tBuOMe gradient from 2% to 40%), separating products **6** (135 mg, 0.25 mmol, y = 13 %) and product **6b** (646 mg, 1.41 mmol, y = 73 %) as anomeric mixtures (ratios 2.5:1 and 1.5:1, respectively). Total yield: 87%. TLC R_f (nHex/tBuOMe: 7/3): 0.15 (**5**), 0.30 (**6**) and 0.15 (**6b**).

6: (1-hydroxy-2,3,4-tri-O-benzyl L-fucopyranosyl) trimethylsilyl acetylene

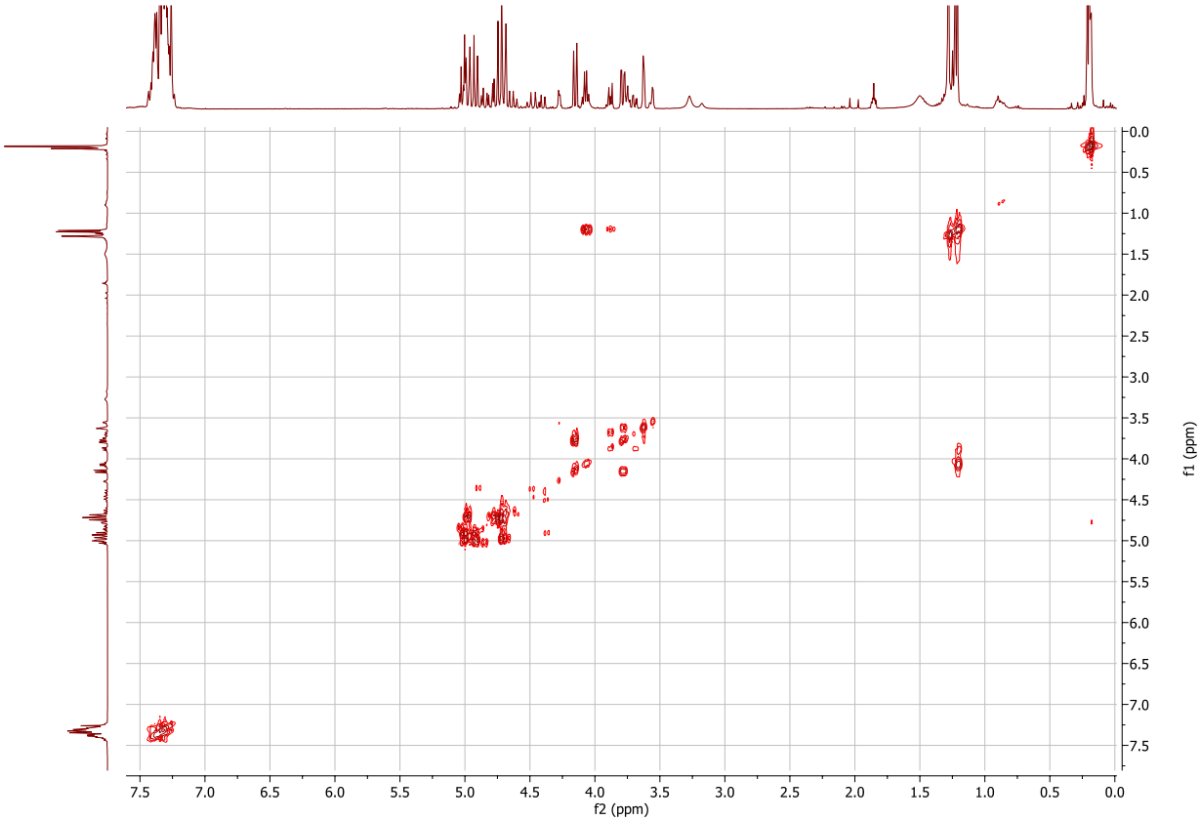
^{1}H NMR (400 MHz, $CDCl_3$):

major-anomer: δ = 7.43 - 7.24 (mult., 15H, *Ar*), 5.03 - 4.68 (mult., 6H, *CH_2-OBn*), 4.15 (d, J_{2-3}= 9.8 Hz, 1H, *H-2*), 4.07 (dq, J_{5-4} = 1.3 Hz, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 3.78 (dd, J_{3-2} = 9.7 Hz, J_{3-4} = 2.9 Hz, 1H, *H-3*), 3.62 (dd, J_{4-3} = 2.9 Hz, J_{4-5} = 1.3 Hz, 1H, *H-4*), 3.27 (bs, 1H, *OH*), 1.22 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*), 0.21 - 0.18 (m, 9H, *Si-CH_3*).

minor-anomer: δ = 7.43 - 7.24 (mult., 15H, *Ar*), 5.03 - 4.38 (mult., 6H, *CH_2-OBn*), 3.88 (mult., J_{2-3}= 10.0 Hz, J_{5-4} = 1.2 Hz, J_{5-CH3} = 6.5 Hz, 2H, *H-2* + *H-5*), 3.69 (dd, J_{3-2} = 10.0 Hz, J_{3-4} = 2.9 Hz, 1H, *H-3*), 3.56 (dd, J_{4-3} = 2.9 Hz, J_{4-5} = 1.3 Hz, 1H, *H-4*), 3.18 (bs, 1H, *OH*), 1.24 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*), 0.21 - 0.18 (m, 9H, *Si-CH_3*).



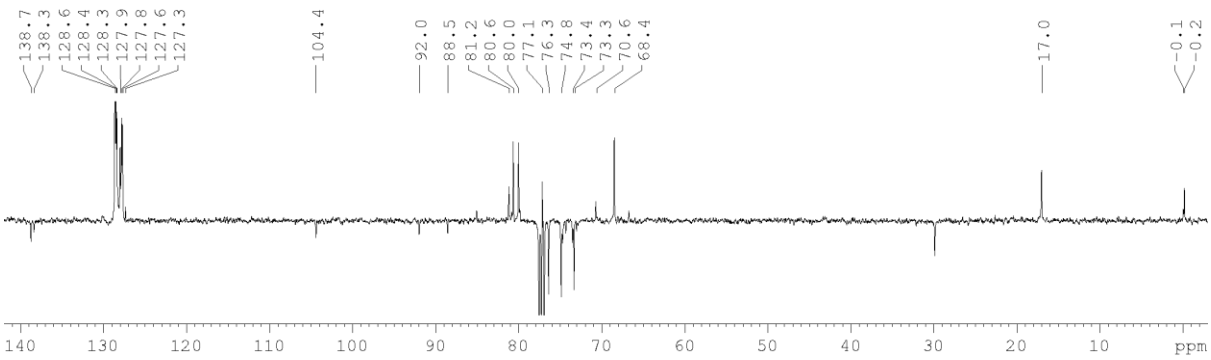
COSY:



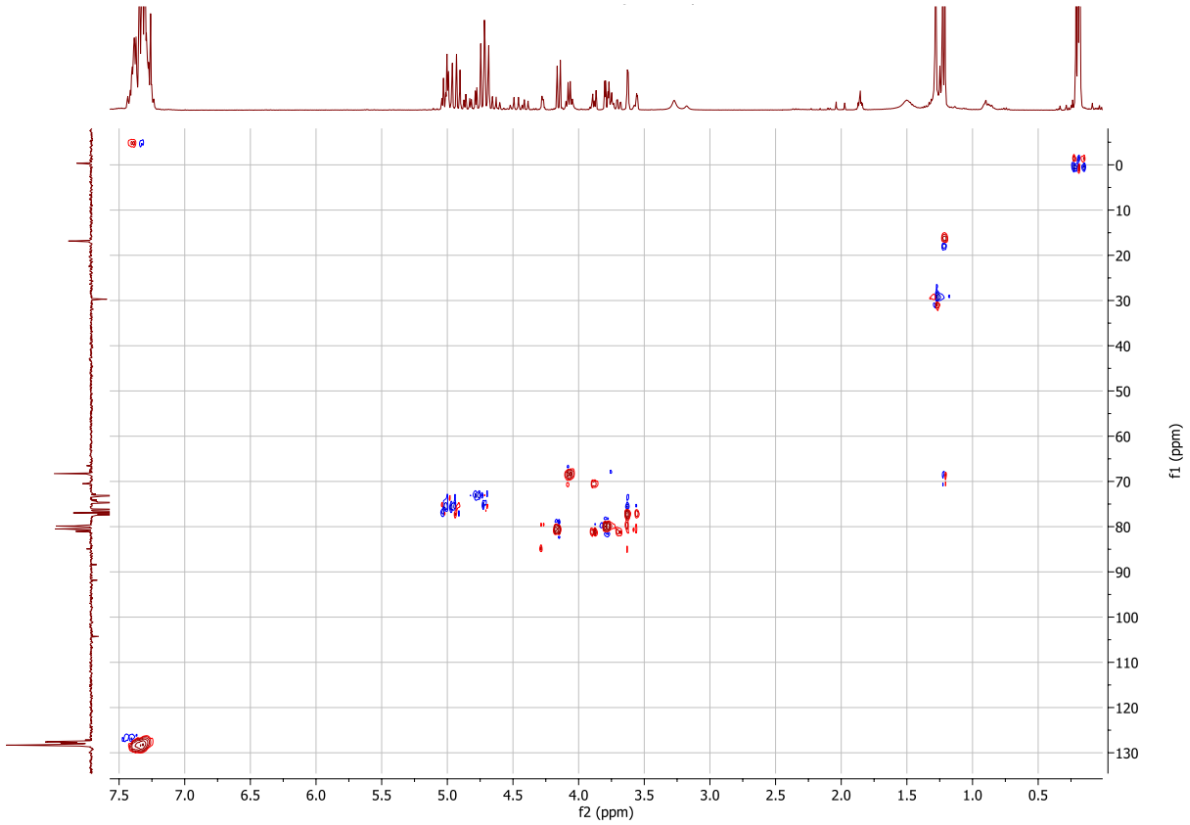
^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

major-anomer: δ = 138.7, 138.3 (*C Ar*), 128.6 - 127.7 (*CH Ar*), 104.4 (*C1*), 92.0 (*C-Si*), 88.5 (-*C≡*), 80.7 (*C2*), 80.0 (*C3*), 77.1 (*C4*), 76.3, 74.8, 73.3 (CH_2 *Ar*), 68.4 (*C5*), 17.0 (*C6*), -0.2 (CH_3-*Si*).

minor-anomer: δ = 81.2 (*C2* + *C3*), 77.1 (*C4*), 70.6 (*C5*), 74.6, 74.3, 73.4 (CH_2 *Ar*), 17.0 (*C6*), -0.1 (CH_3-*Si*).



HSQC:

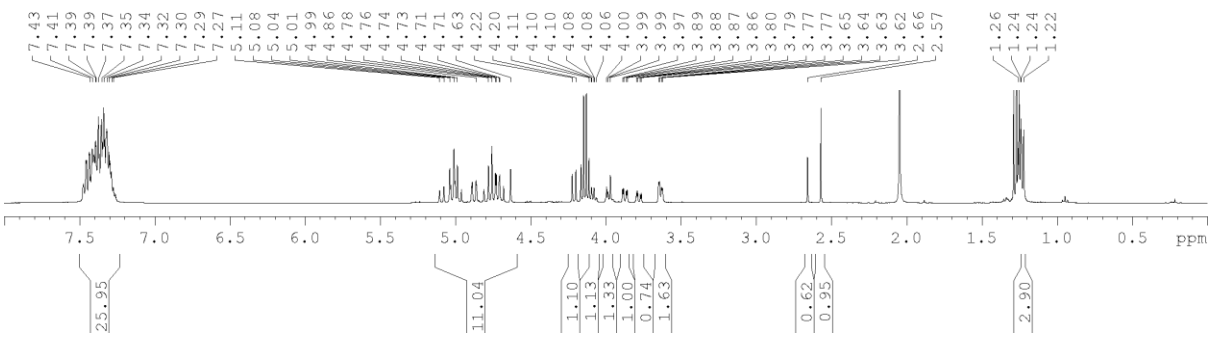


6b: (1-hydroxy-2,3,4-tri-O-benzyl L-fucopyranosyl) acetylene

^{1}H NMR (400 MHz, $CDCl_3$):

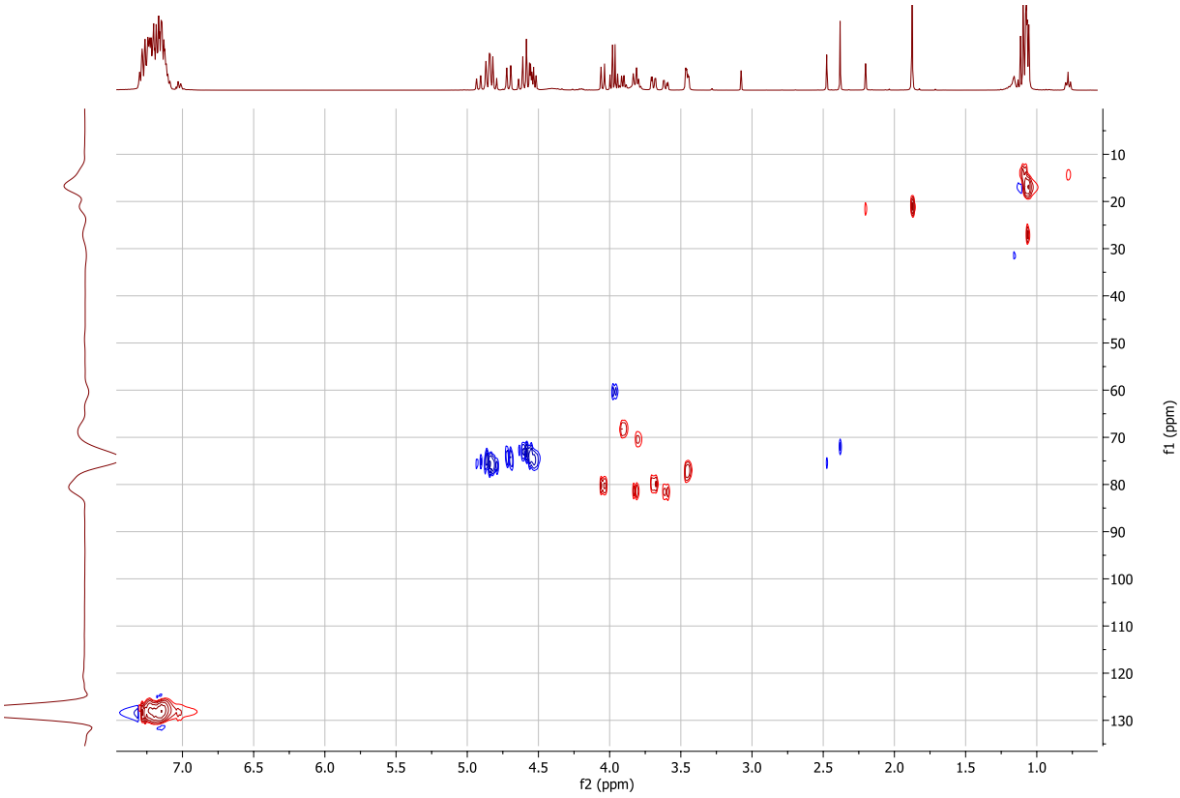
major-anomer: δ = 7.47 - 7.27 (mult., 15H, *Ar*), 5.11 - 4.63 (mult., 6H, *CH_2-OBn*), 4.21 (d, $J_{2\text{-}3}$= 9.8 Hz, 1H, *H-2*), 4.09 (dq, $J_{5\text{-}4}$ = 1.2 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.87 (dd, $J_{3\text{-}2}$ = 9.8 Hz, $J_{3\text{-}4}$ = 2.8 Hz, 1H, *H-3*), 3.65 (dd, $J_{4\text{-}3}$ = 2.9 Hz, $J_{4\text{-}5}$ = 1.2 Hz, 1H, *H-4*), 2.57 (s, 1H, *≡CH*), 1.23 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*).

minor-anomer: δ = 7.47 - 7.27 (mult., 15H, *Ar*), 5.11 - 4.63 (mult., 6H, *CH_2-OBn*), 3.98 (m, $J_{2\text{-}3}$= 10.0 Hz, $J_{5\text{-}4}$ = 1.2 Hz, $J_{5\text{-}CH3}$ = 6.4 Hz, 1H, *H-2* + *H-5*), 3.78 (dd, $J_{3\text{-}2}$ = 10.0 Hz, $J_{3\text{-}4}$ = 2.9 Hz, 1H, *H-3*), 3.63 (dd, $J_{4\text{-}3}$ = 2.9 Hz, $J_{4\text{-}5}$ = 1.2 Hz, 1H, *H-4*), 2.66 (s, 1H, *≡CH*), 1.25 (d, $J_{CH3\text{-}5}$ = 6.4 Hz, 3H, *CH_3*).



^{13}C NMR shifts extrapolated from the HSQC experiment:

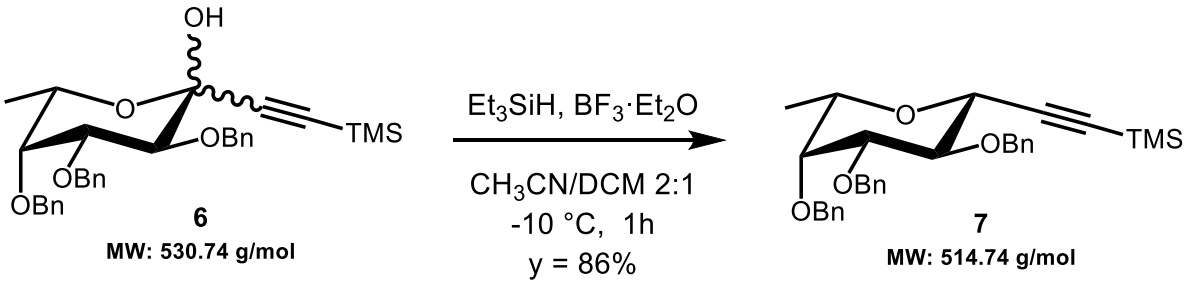
HSQC:



major-anomer: δ = 128.0 (*CH Ar*), 80.2 (*C2*), 80.0 (*C3*), 77.2 (*C4*), 74.3 (*CH_2 Ar*), 72.1 (≡*CH*) 68.3 (*C5*), 16.8 (*C6*).

minor-anomer: δ = 128.0 (*CH Ar*), 81.6 (*C3*), 81.4 (*C2*), 77.2 (*C4*), 75.5 (≡*CH*), 74.3 (*CH_2 Ar*), 70.5 (*C5*), 16.8 (*C6*).

(7) Synthesis and characterization of **(2,3,4-tri-O-benzyl β-L-fucopyranosyl) trimethylsilyl acetylene (7)** following the procedure of Lowary and co-workers:[181]

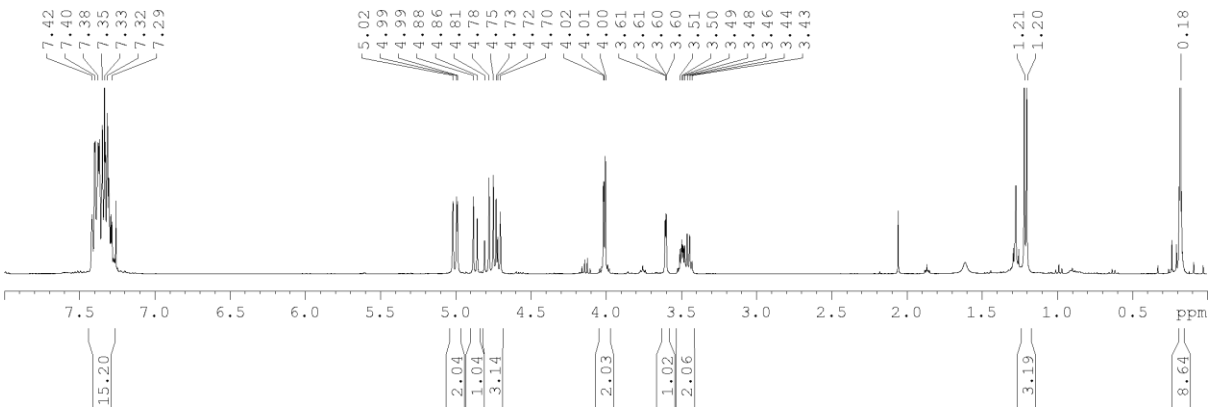


A solution of **1-hydroxy-2,3,4-tri-O-benzyl L-fucopyranosyl) trimethylsilyl acetylene 6** (213 mg, 0.40 mmol, 1 eq) dissolved in CH_3CN/DCM (ratio 2:1 - 5.35 mL, 2.67 mL) was cooled to -10 °C under Ar atmosphere. Et_3SiH (256 µL, 1.61 mmol, 4.0 eq), then $BF_3{\cdot}Et_2O$ (248 µL, 2.01

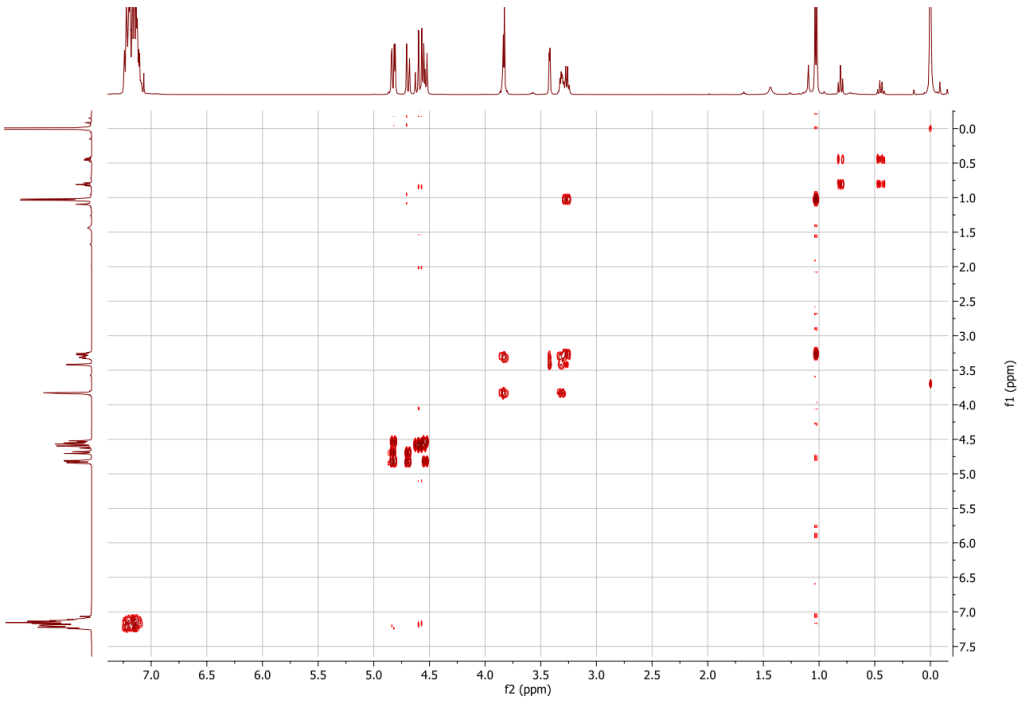
mmol, 5.0 eq) were added to the solution and left to stir at -10 °C for 1h until TLC showed completion, before returning to room temperature. The reaction mixture was quenched with a few drops of Et_3N and diluted in EtOAc. The organic phase was washed with water and brine, then was dried over Na_2SO_4. The crude product was purified by automatic chromatography (Biotage SNAP 10: nHex/tBuOMe from 0 % to 50 %) affording product **7** (178 mg, 0.35 mmol, y = 86 %). TLC R_f (nHex/tBuOMe: 7/3): 0.70.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.40 - 7.29 (mult., 15H, *Ar*), 5.02 - 4.99 (m, 2H, *CH_2-OBn*), 4.87 (d, 1H, *CH_2-OBn*), 4.81 - 4.70 (mult., 3H, *CH_2-OBn*), 4.02 - 4.00 (mult., 2H, *H-1* + *H-2*), 3.60 (dd, $J_{4\text{-}3}$ = 2.9 Hz, $J_{4\text{-}5}$ = 1.1 Hz, 1H, *H-4*), 3.49 (m, 1H, *H-3*), 3.45 (dq, $J_{5\text{-}4}$ = 1.1 Hz, $J_{5\text{-}CH3}$ = 6.4 Hz, 1H, *H-5*), 1.21 (d, $J_{CH3\text{-}5}$ = 6.4 Hz, 3H, *CH_3*), 0.18 (m, 9H, *Si-CH_3*).

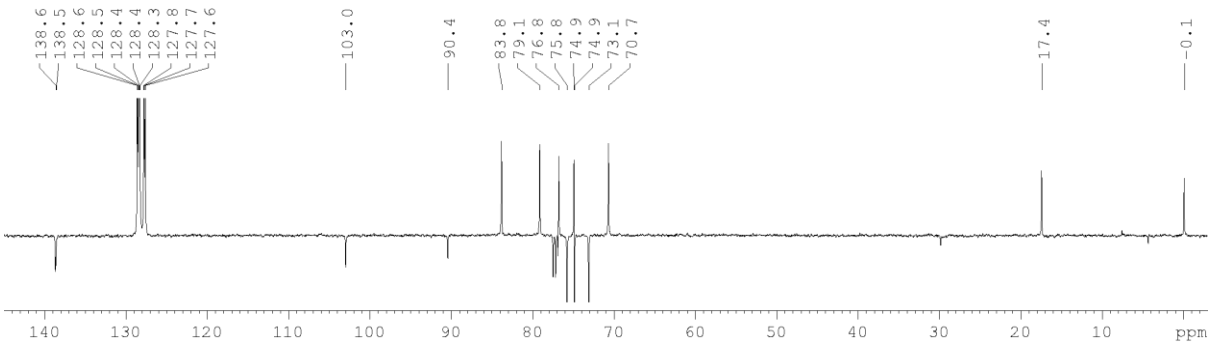


COSY:

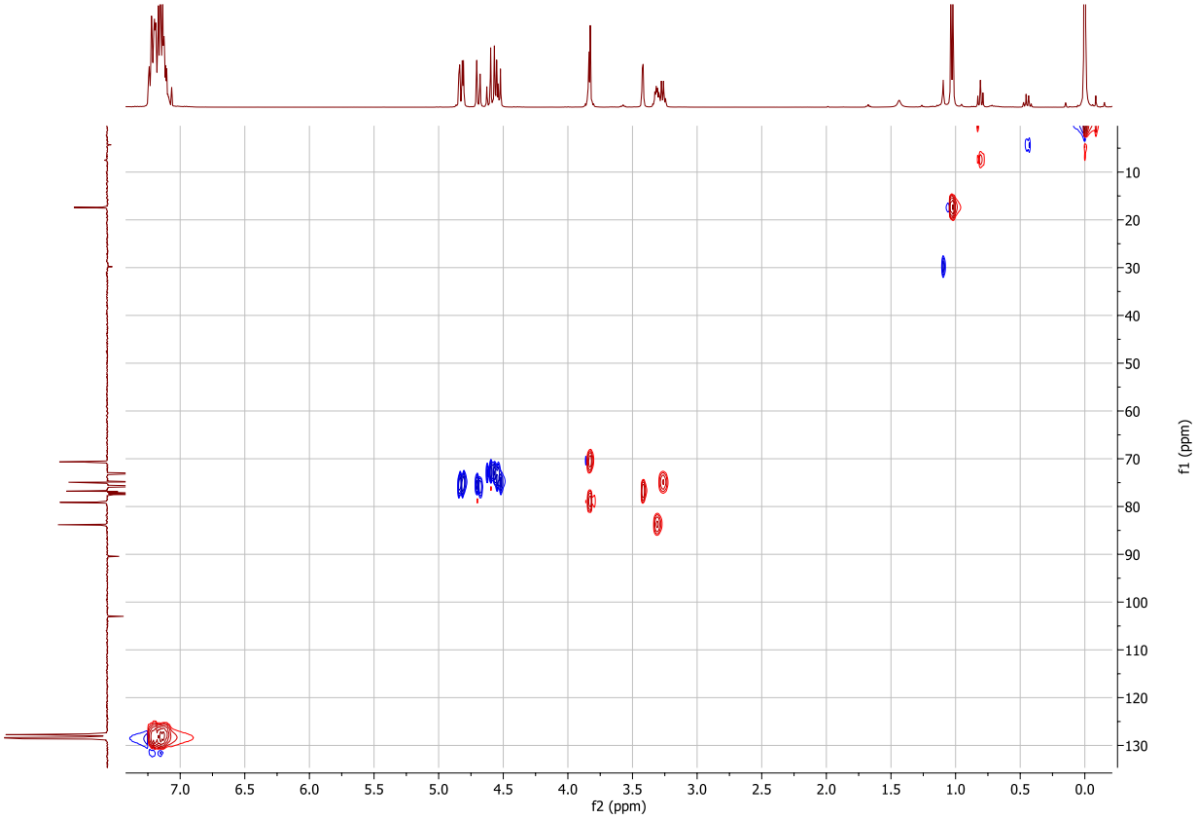


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

δ = 138.6, 138.5 (*C Ar*), 128.6 – 127.6 (*CH Ar*), 103.0 (-*C*≡), 90.4 (*C-Si*), 83.8 (*C3*), 79.1 (*C2*), 76.8 (*C4*), 75.8, 74.9, 73.1 (CH_2 *Ar*), 74.9 (*C5*), 70.7 (*C1*), 17.4 (*C6*), -0.1 (CH_3-*Si*).

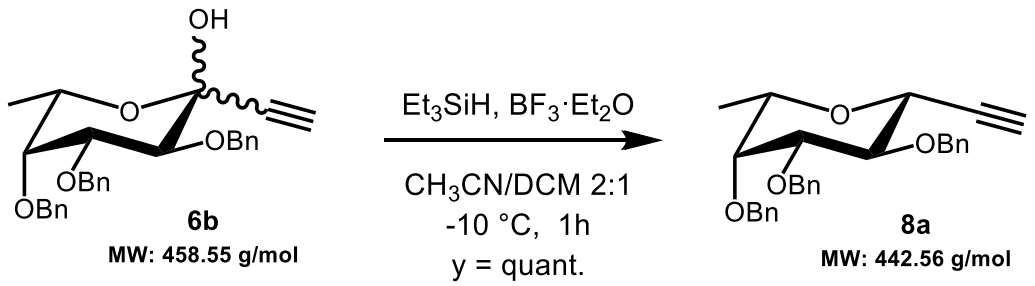


HSQC:



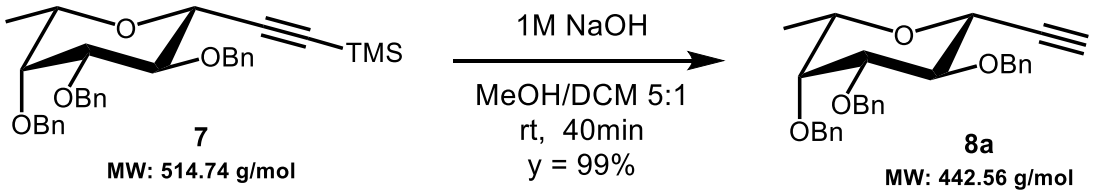
(8a) Synthesis and characterization of **(2,3,4-tri-O-benzyl β-L-fucopyranosyl) acetylene (8a)**:

From **6b**:



(1-hydroxy-2,3,4-tri-O-benzyl L-fucopyranosyl) acetylene 6b (1.583 g, 3.45 mmol) was reduced following the procedure described above to afford crude **8a** (2.606 g, quantitative), used for the following step without further purification.

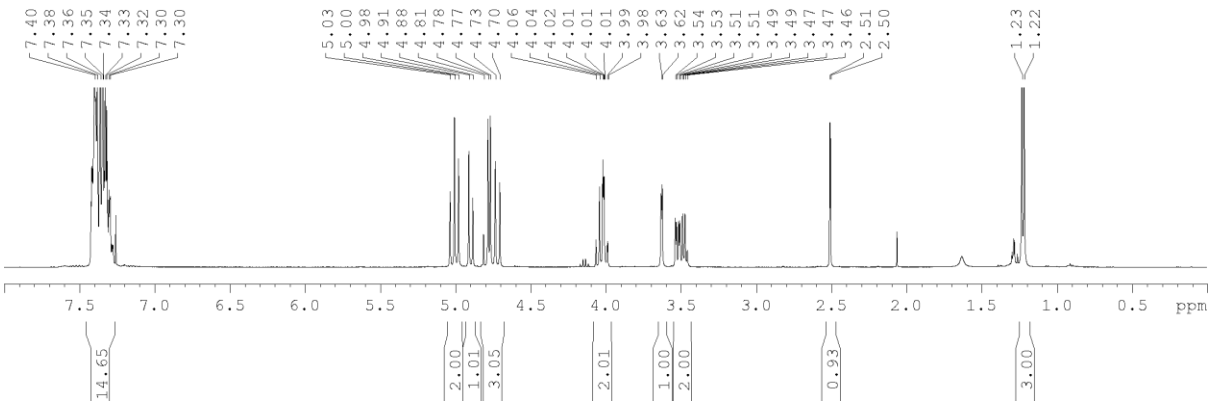
From **7**, following the procedure of Dondoni and co-workers:[192]



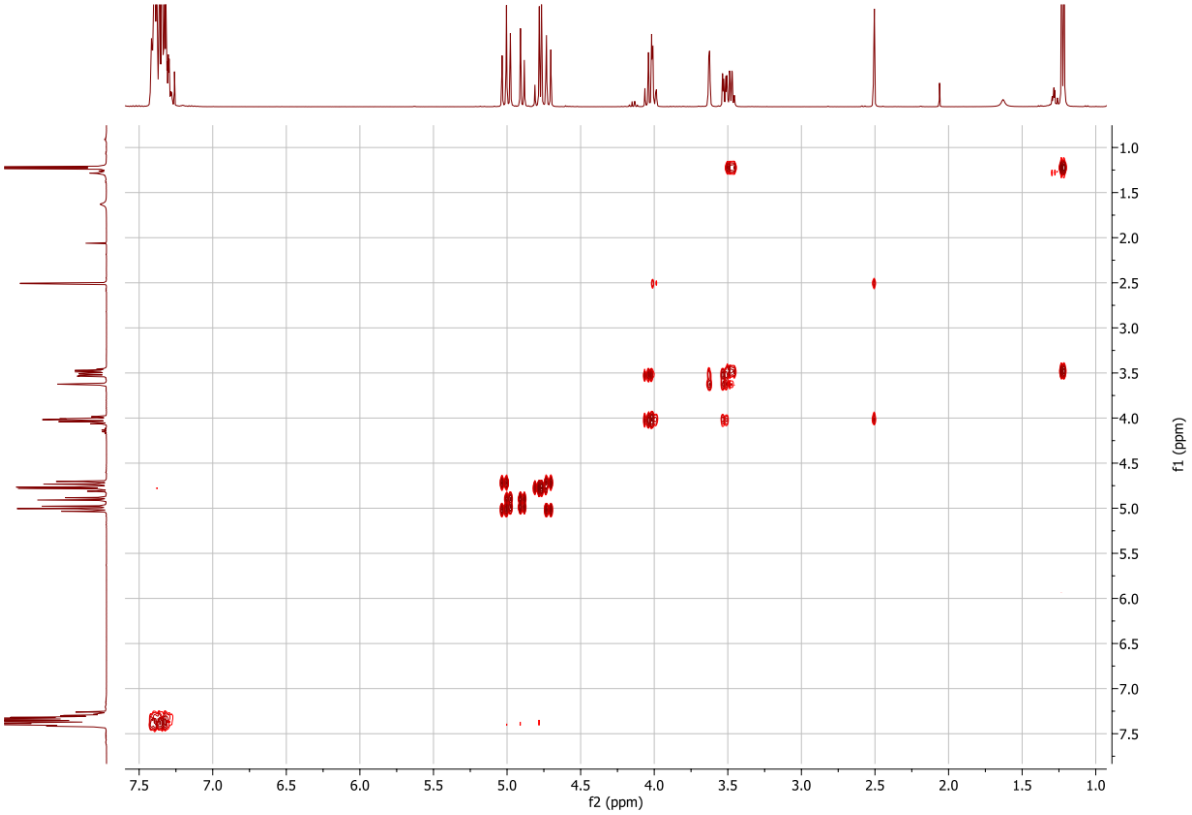
To a solution of **(2,3,4-tri-O-benzyl β-L-fucopyranosyl) trimethylsilyl acetylene 7** (122 mg, 0.24 mmol, 1 eq) dissolved in MeOH/DCM (ratio 5:1 - 7.9 mL) was added a 1M solution of NaOH (415 µL, 42 mmol, 1.75 eq). The reaction mixture was stirred at room temperature for 40 min until TLC showed completion, before it was quenched with a few drops of a 2 N HCl solution. The solvents were evaporated and the resulting crude was dissolved in water and extracted with EtOAc. The organic phase was washed with water and dried over Na_2SO_4. The crude product was purified by flash chromatography (nHex/tBuOMe 7:3) affording product **8a** (104 mg, 0.24 mmol, y = 99 %). TLC R_f (nHex/tBuOMe: 7/3): 0.54. $[\alpha]_D^{19}$ = -2.5 ($CHCl_3$, c 0.9).

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.42 - 7.28 (mult., 15H, *Ar*), 5.03 - 4.98 (m, 2H, *CH_2-OBn*), 4.90 (d, 1H, *CH_2-OBn*), 4.81 - 4.70 (mult., 3H, *CH_2-OBn*), 4.02 (mult., 2H, *H-1* + *H-2*), 3.63 (dd, J_{4-3} = 2.9 Hz, J_{4-5} = 1.1 Hz, 1H, *H-4*), 3.52 (dd, J_{3-2} = 8.4 Hz, J_{3-4} = 2.9 Hz, 1H, *H-3*), 3.48 (dq, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.4 Hz, 1H, *H-5*), 2.51 (d, J = 1.90 Hz, *≡CH*), 1.22 (d, J_{CH3-5} = 6.4 Hz, 3H, *CH_3*).

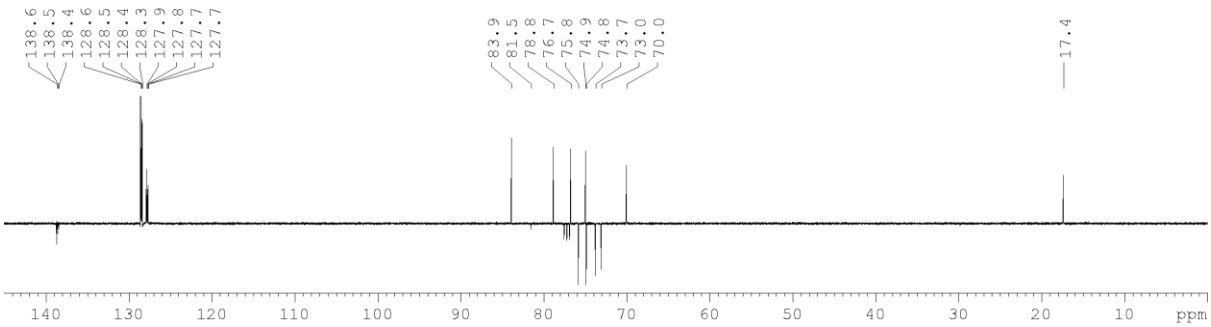


COSY:

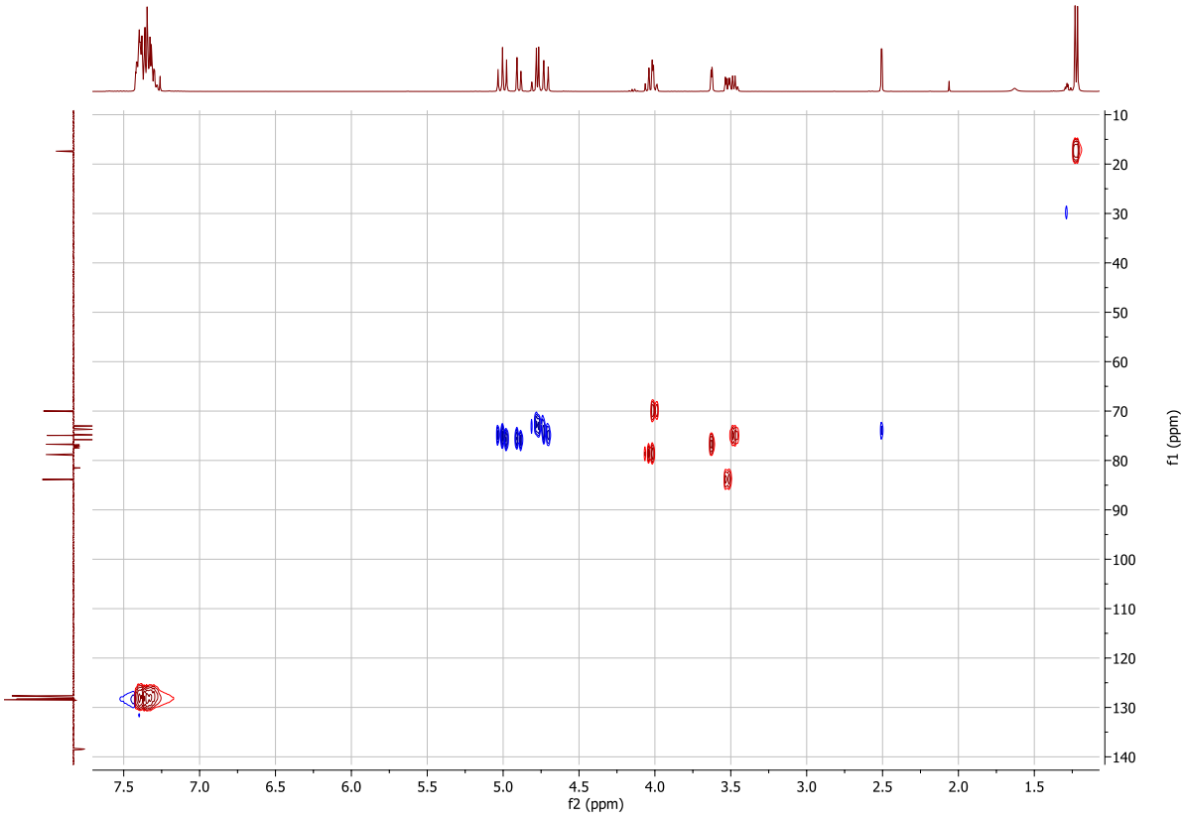


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

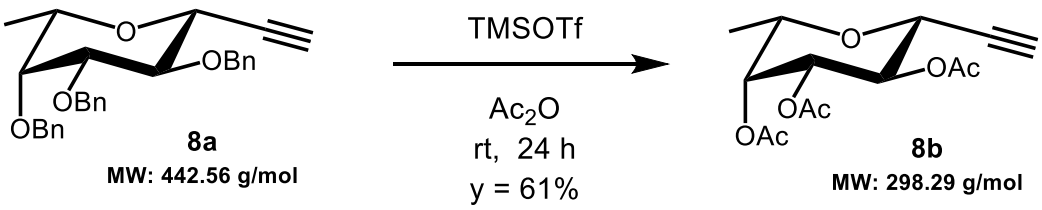
δ = 138.6, 138.5, 138.4 (*C Ar*), 128.6 – 127.7 (*CH Ar*), 83.9 (*C3*), 81.5 (-*C*≡), 78.8 (*C2*), 76.7 (*C4*), 75.8, 74.8, 73.0 (CH_2 *Ar*), 74.9 (*C5*), 73.7 (≡*CH*), 70.0 (*C1*), 17.4 (*C6*).



HSQC:



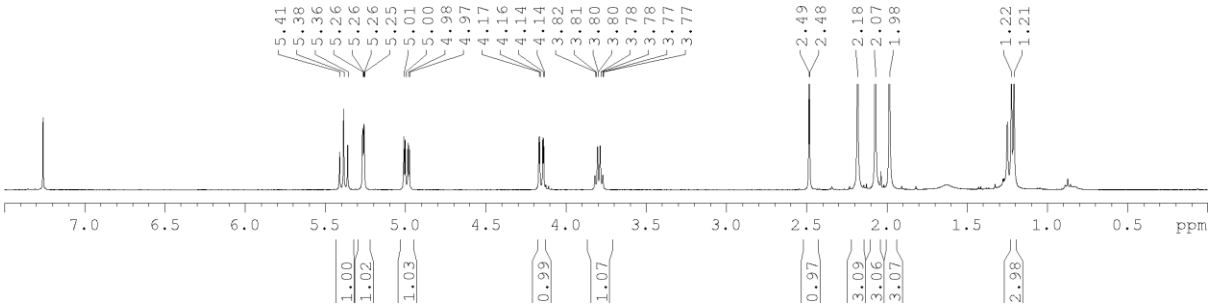
(8b) Synthesis and characterization of **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) acetylene (8b)** following the procedure of Alzeer and co-workers:[180]



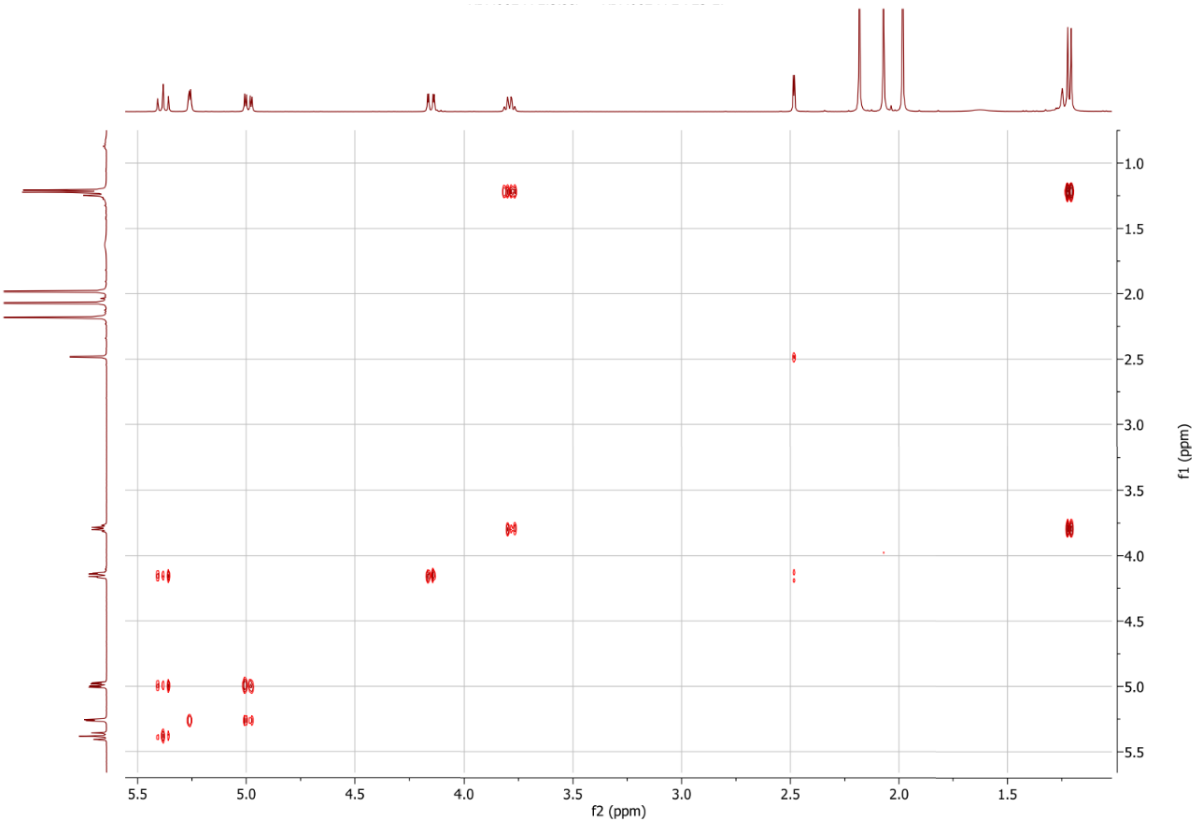
To a solution of **(2,3,4-tri-O-benzyl β-L-fucopyranosyl) acetylene 8a** (856 mg, 1.93 mmol, 1 eq) dissolved in Ac_2O (38 mL) was added a solution of TMSOTf (3.2 mL, 17.64mmol, 9.1 eq) under Ar atmosphere. The brown reaction mixture was stirred at room temperature for 24h until TLC showed completion, before it was transferred to a separatory funnel and carefully quenched with a $NaHCO_3$ saturated aqueous solution. The mixture was extracted with EtOAc, washed with water and brine, then dried over Na_2SO_4. The crude product was purified by automatic chromatography (Biotage SNAP 100: nHex/EtOAc from 5 % to 70 %) affording product **8b** (354 mg, 1.19 mmol, y = 61 %). TLC R_f (nHex/EtOAc: 7/3): 0.31. $[\alpha]_D^{19}$ = -30 ($CHCl_3$, c 1). MS (ESI) calculated for $C_{14}H_{18}O_7$ $[M + H]^+$ *m/z*: 299.11; found: 299.05.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 5.38 (dd, , J_{2-1} = J_{2-3} = 10.0 Hz, 1H, *H-2*), 5.26 (dd, J_{4-3} = 3.4 Hz, J_{4-5} = 1.2 Hz, 1H, *H-4*), 4.99 (dd, J_{3-2} = 10.2 Hz, J_{3-4} = 3.4 Hz, 1H, *H-3*), 4.15 (dd, J_{1-2} = 9.9 Hz, J_{alkyne} = 2.2 Hz, 1H, *H-1*), 3.79 (dq, J_{5-4} = 1.2 Hz, J_{5-CH3} = 6.4 Hz, 1H, *H-5*), 2.48 (d, J_{alkyne} = 2.2 Hz, 1H, *≡CH*), 2.18 (s, 3H, *OAc*), 2.07 (s, 3H, *OAc*), 1.98 (s, 3H, *OAc*), 1.21 (d, J_{CH3-5} = 6.4 Hz, 3H, *CH_3*).

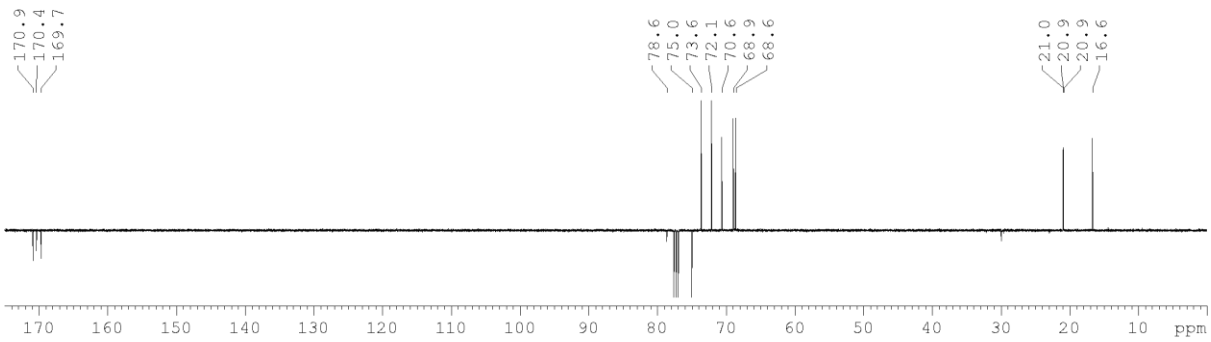


COSY:

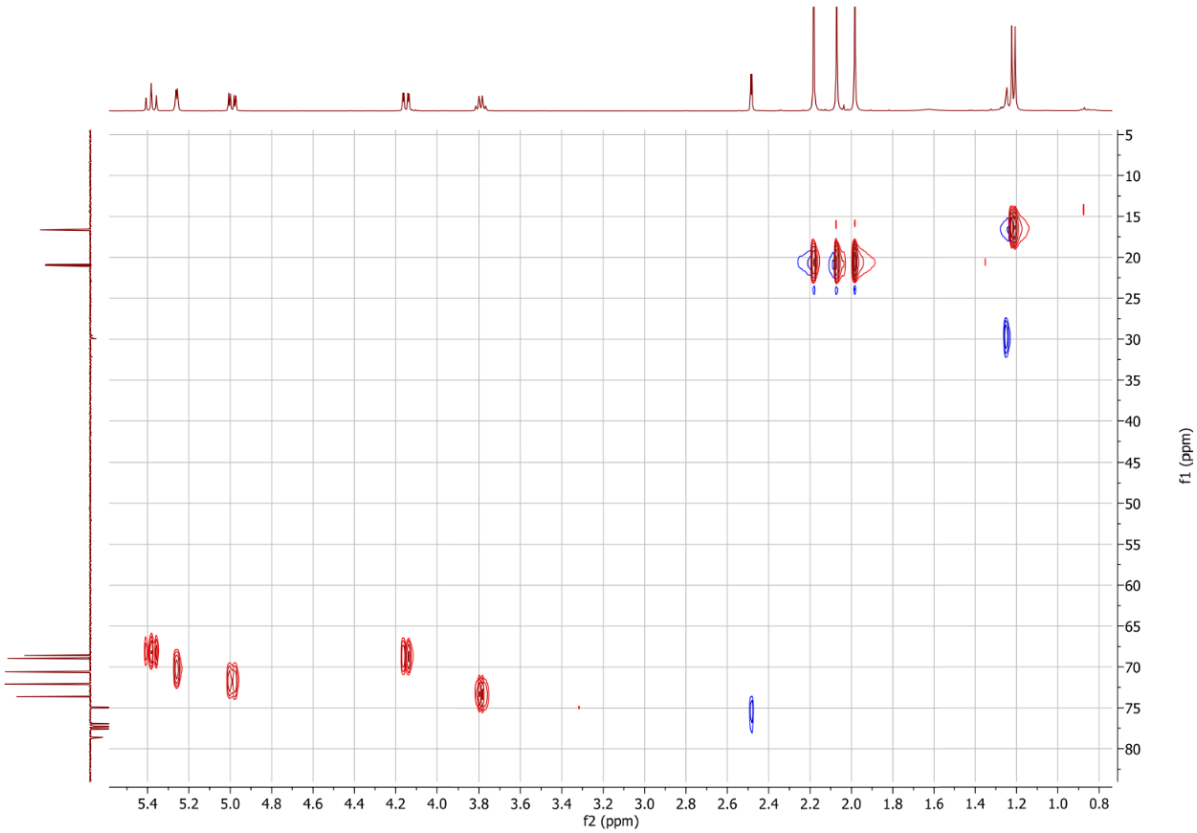


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

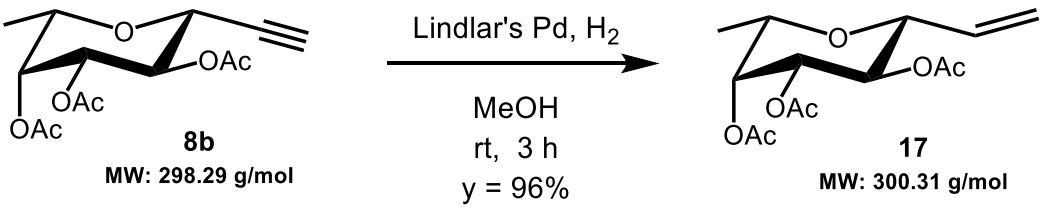
δ = 170.9, 170.4, 169.7 (*C=O*), 78.6 (*-C≡*), 75.0 (*≡CH*), 73.6 (*C5*), 72.1 (*C3*), 70.6 (*C4*), 68.9 (*C1*), 68.6 (*C2*), 21.0, 20.9, 20.9 (*CH_3 OAc*), 16.6 (*C6*).



HSQC:



(17) Synthesis and characterization of **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) ethene (17)** following the procedure of Rouzier and co-workers:[190]

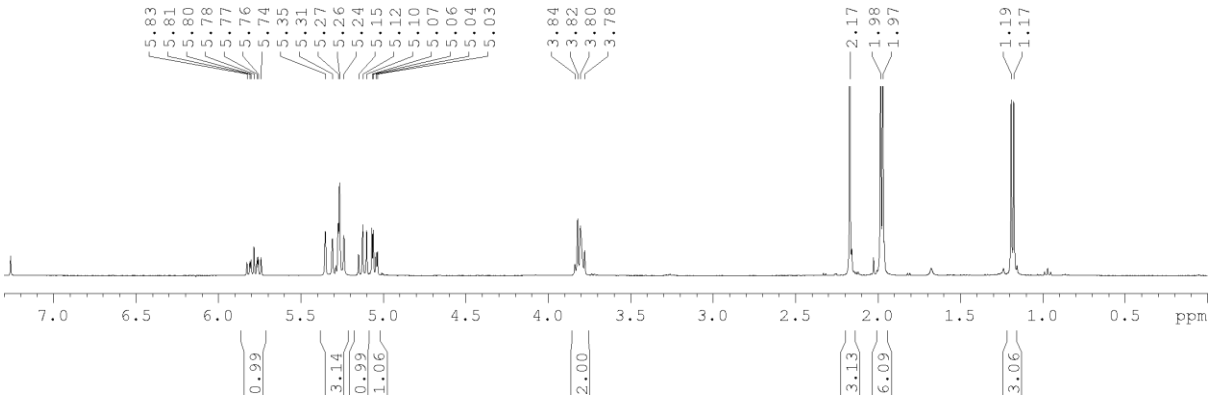


To a solution of **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) acetylene (8b)** (54 mg, 0.181 mmol, 1 eq) dissolved in MeOH (1.8 mL) was added Lindlar's catalyst (4mg, 0.019mmol, 0.2 eq). The mixture was put under H_2 atmosphere (1atm) and stirred a room temperature for 3 h before filtering and concentrating. The crude product **17** (48 mg, 0.160 mmol, y = 89%) was used for

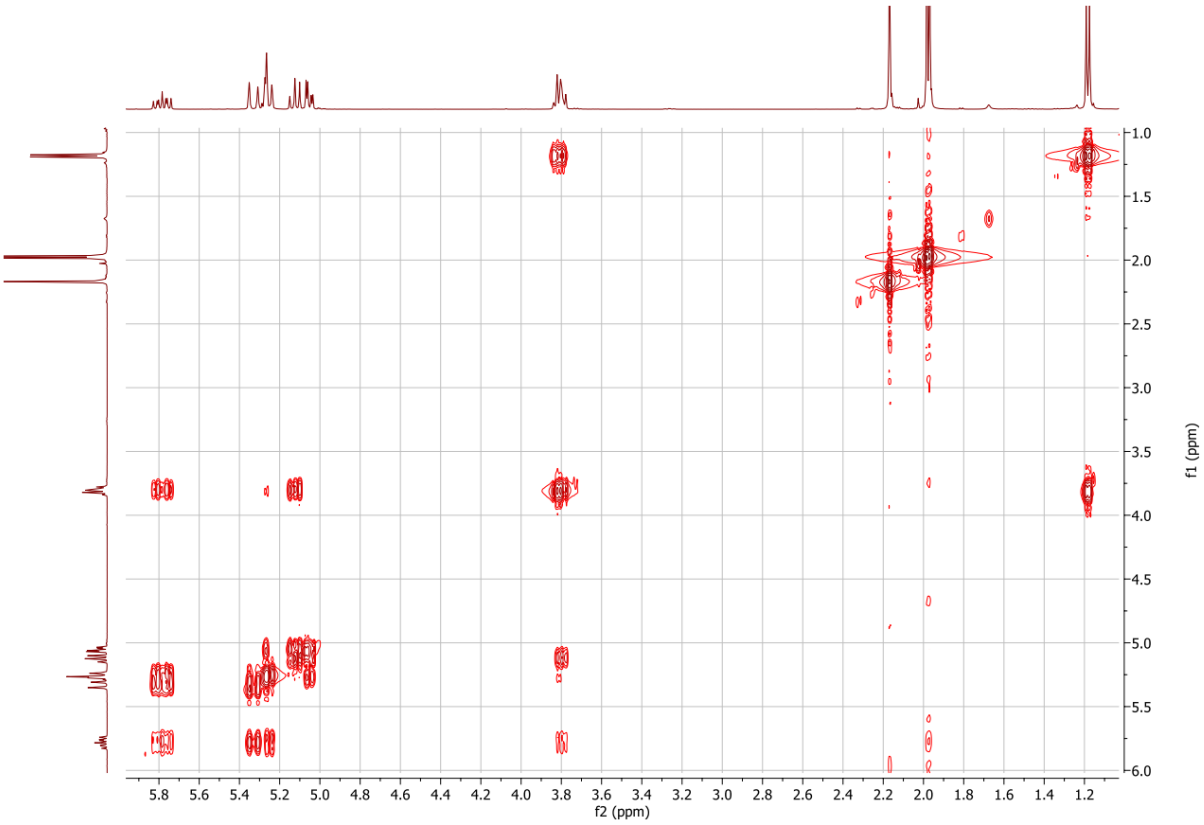
the next step without further purification. TLC R_f (nHex/EtOAc: 7/3): 0.45. MS (ESI) calculated for $C_{14}H_{20}O_7$ $[M + Na]^+$ *m/z*: 323.11; found: 323.27.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 5.78 (ddd, J_{trans} = 17.5 Hz, J_{cis} = 10.4 Hz, J_{CH-1} = 7.4 Hz, 1H, *CH*=*CH*$_2$), 5.29 (dd, J_{trans} = 17.2 Hz, J_{gem} = 1.2 Hz, 2H, *CH*=*CH*$_2$), 5.27 (dd, J_{4-3} = 3.4 Hz, J_{4-5} = 1.2 Hz, 1H,*H-4*), 5.12 (dd, J_{2-1} = 9.4 Hz, J_{2-3} = 10.2 Hz, 1H, *H-2*), 5.05 (dd, J_{3-2} = 10.2 Hz, J_{3-4} = 3.4 Hz, 1H, *H-3*), 3.81 (mult., J_{5-4} = 1.2 Hz, J_{5-CH3} = 6.5 Hz, J_{1-2} = 9.4 Hz, J_{1-CH} = 7.4 Hz, 2H, *H-4* + *H-1*), 2.17 (s, 3H, *OAc*), 1.98 (s, 3H, *OAc*), 1.97 (s, 3H, *OAc*), 1.18 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH*$_3$).

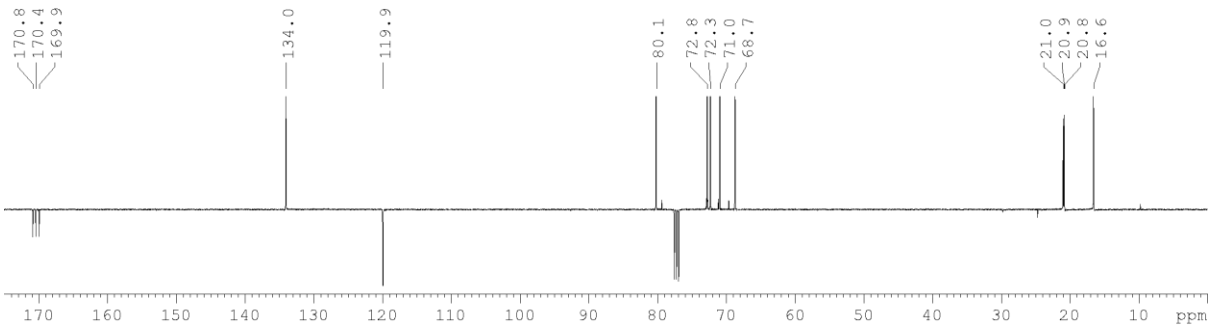


COSY:

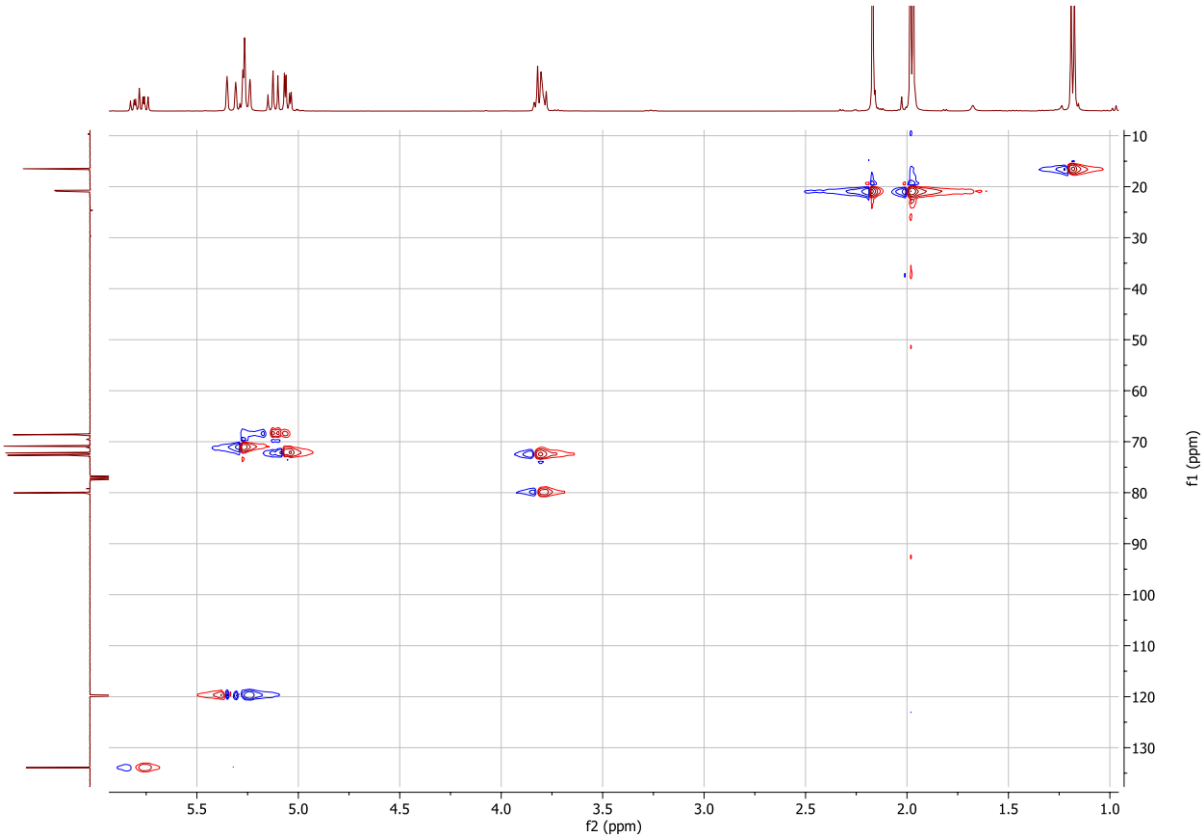


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

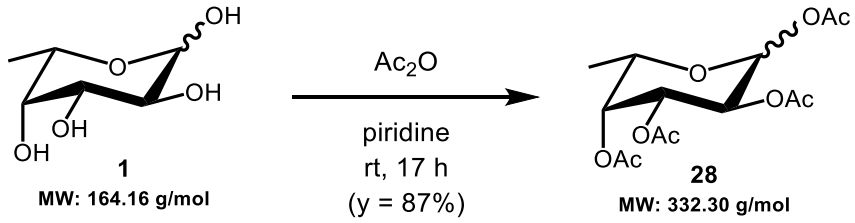
δ = 170.8, 170.4, 169.9 (*C=O*), 134.0 (*CH=CH₂*), 119.9(*CH=CH₂*), 80.1 (*C1*), 72.8 (*C5*), 72.3 (*C3*), 71.0 (*C4*), 68.7 (*C2*), 21.0, 20.9, 20.8 (*CH₃ OAc*), 16.6 (*C6*).



HSQC:



(28) Synthesis and characterization of **1,2,3,4-tetra-O-acetyl L-fucopyranoside (28)** following the procedure of Duléry and co-workers:[220]



To a solution of L-fucose **1** (150 mg, 0.91 mmol, 1 eq) in pyridine (2 mL) was added acetic anhydride (2 mL). The reaction mixture was stirred at room temperature for 17 h until TLC showed completion, before being co-evaporated with toluene. The crude **28** (264 mg, 0.79 mmol, y = 87 %) was composed of the α-pyranoside (80 %) and β-pyranoside (20 %) species with traces of furanoside, as observed by ^{1}H NMR (α/β anomeric protons: 6.34/5.68 ppm). This mixture was directly used for the next step. TLC R_f (EtOAc/MeOH: 95/5): 0.90.

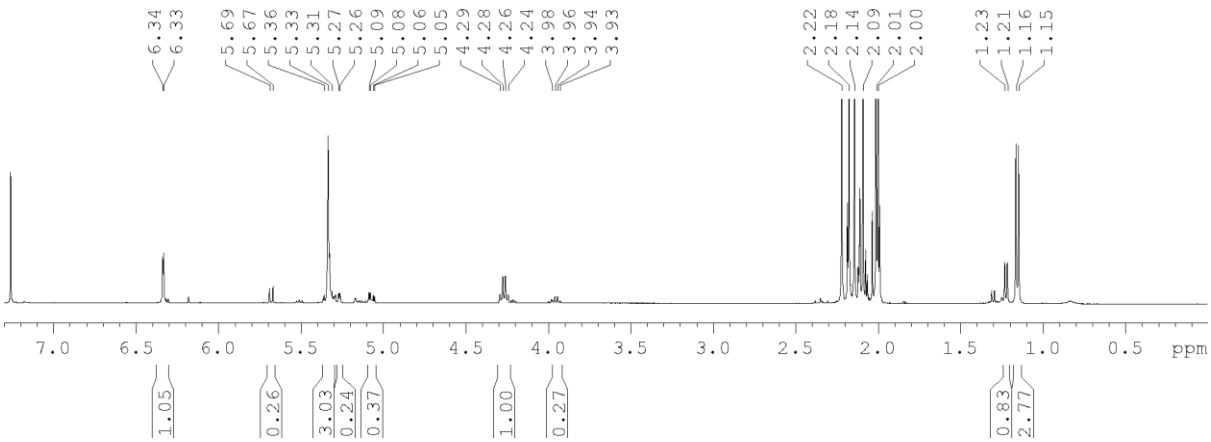
^{1}H NMR (400 MHz, $CDCl_3$): Crude mixture of α-pyranoside and β-pyranoside (80:20)

α-anomer:

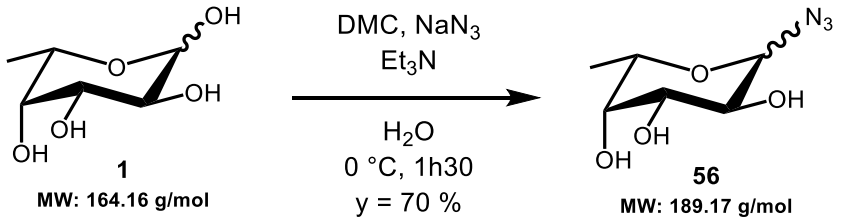
δ = 6.34 (d, $J_{1\text{-}2}$ = 2.7 Hz, 1H, *H-1*), 5.36 - 5.31 (mult., 3H, *H-2 + H-3 + H-4*), 4.27 (q, $J_{5\text{-}CH3}$ = 6.4 Hz, 1H, *H-5*), 2.22 - 2.00 (s, 3H, *OAc*), 1.15 (d, $J_{CH3\text{-}5}$ = 6.4 Hz, 3H, *CH_3*). In accordance with published data.[220]

β-anomer:

δ = 5.68 (d, $J_{1\text{-}2}$ = 8.3 Hz, 1H, *H-1*), 5.27 (dd, $J_{4\text{-}3}$ = 3.4 Hz, $J_{4\text{-}5}$ = 1.1 Hz, 1H, *H-4*), 3.95 (m, 1H, *H-5*), 2.22 - 2.00 (s, 3H, *OAc*), 1.22 (d, $J_{CH3\text{-}5}$ = 6.3 Hz, 3H, *CH_3*).



(56) Synthesis and characterization of **L-fucopyranosyl azide (56)** following the procedure of Tanaka and co-workers:[198]



A solution of L-fucose **1** (305 mg, 1.86 mmol, 1 eq) in H_2O (7.5 mL) was cooled to 0 °C prior to addition 2-chloro-1,3-dimethylimidazolinium chloride - DMC (926 mg, 5.48 mmol, 2.95 eq) NaN_3 (1.78 g, 27.45 mmol, 14.8 eq). Et_3N (2.3 mL, 16.47 mmol, 8.9 eq) was added and the solution was stirred for 1h30 while returning to room temperature until TLC showed completion. The reaction mixture was concentrated and redissolved in EtOH and filtered. The crude was purified by automatic chromatography (Biotage SNAP 25: CH_2Cl_2/MeOH gradient from 0% to 25%), affording product **56** (244 mg, 1.29 mmol, y = 70 %) as a mixture of β-pyranoside (73 %) and α-pyranoside (27 %), as observed by ^{1}H NMR(α/β anomeric protons: 5.52/4.66 ppm). This mixture was inseparable through reverse phase chromatography and proceeded to the next step. TLC R_f (DCM/MeOH: 5/1): 0.60.

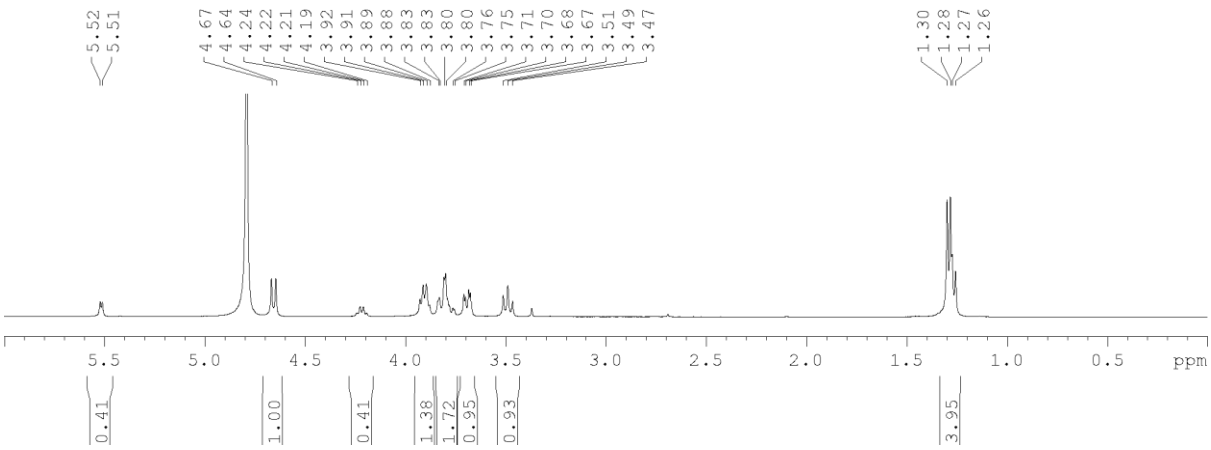
^{1}H NMR (400 MHz, D_2O): Crude mixture of α-pyranoside and β-pyranoside (80:20)

β-anomer:

δ = 4.66 (d, $J_{1\text{-}2}$ = 9.3 Hz, 1H, *H-1*), 3.90 (q, $J_{5\text{-}CH3}$ = 6.6 Hz, 1H, *H-5*), 3.80 (m, 1H, *H-4*), 3.69 (dd, $J_{3\text{-}2}$ = 9.3 Hz, $J_{3\text{-}4}$ = 3.5 Hz, 1H, *H-3*), 3.49 (t, $J_{2\text{-}3}$ = $J_{2\text{-}1}$ = 9.3 Hz, 1H, *H-2*), 1.29 (d, $J_{CH3\text{-}5}$ = 6.6 Hz, 3H, *CH_3*). In accordance with published data.[221]

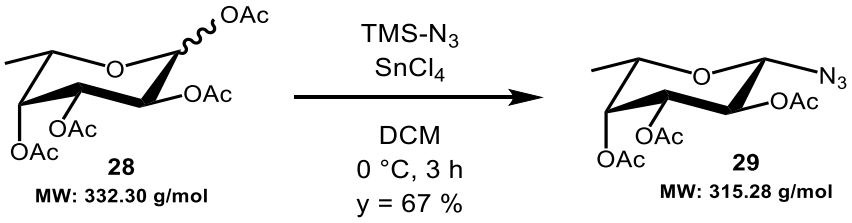
α-anomer:

δ = 5.52 (d, $J_{1\text{-}2}$ = 4.5 Hz, 1H, *H-1*), 4.22 (q, $J_{5\text{-}CH3}$ = 6.6 Hz, 1H, *H-5*), 3.92 - 3.75 (m, 3H, *H-2* + *H-3* + *H-4*), 1.27 (d, $J_{CH3\text{-}5}$ = 6.6 Hz, 3H, *CH_3*).



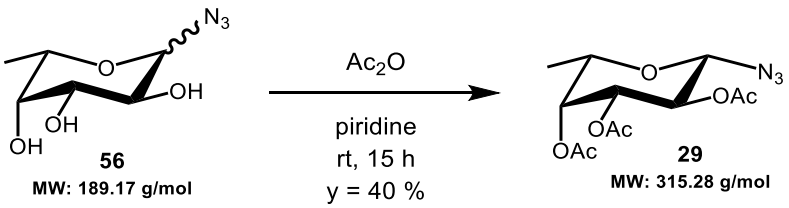
(29) Synthesis and characterization of **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide (29)**

From **28**, following the procedure of Palomo and co-workers:[222]



1,2,3,4-tetra-O-acetyl L-fucopyranoside 28 (264 mg, 0.79 mmol, 1eq, α/β ratio: 80/20) was dissolved in dry DCM (4 mL) and cooled to 0°C under N_2 atmosphere. TMS-N_3 (136 µL, 1.03 mmol, 1.3 eq) and $SnCl_4$ (46 µL, 0.40 mmol, 0.5 eq) were added and the reaction mixture was stirred for 3 h while returning to room temperature until TLC showed completion. The reaction was diluted with DCM and washed with $NaHCO_3$ saturated aqueous solution and water. The organic phase was dried over Na_2SO_4 and concentrated, affording a crude anomeric mixture (α/β ratio: 9/91), as observed by ^{1}H NMR (α/β anomeric protons: 5.61/4.58 ppm). The crude was product was purified by flash chromatography (nHex/EtOAc 8:2) affording product **29** (192 mg, 0.61 mmol, y = 67 % over two steps). TLC R_f (nHex/EtOAc: 3/1): 0.46.

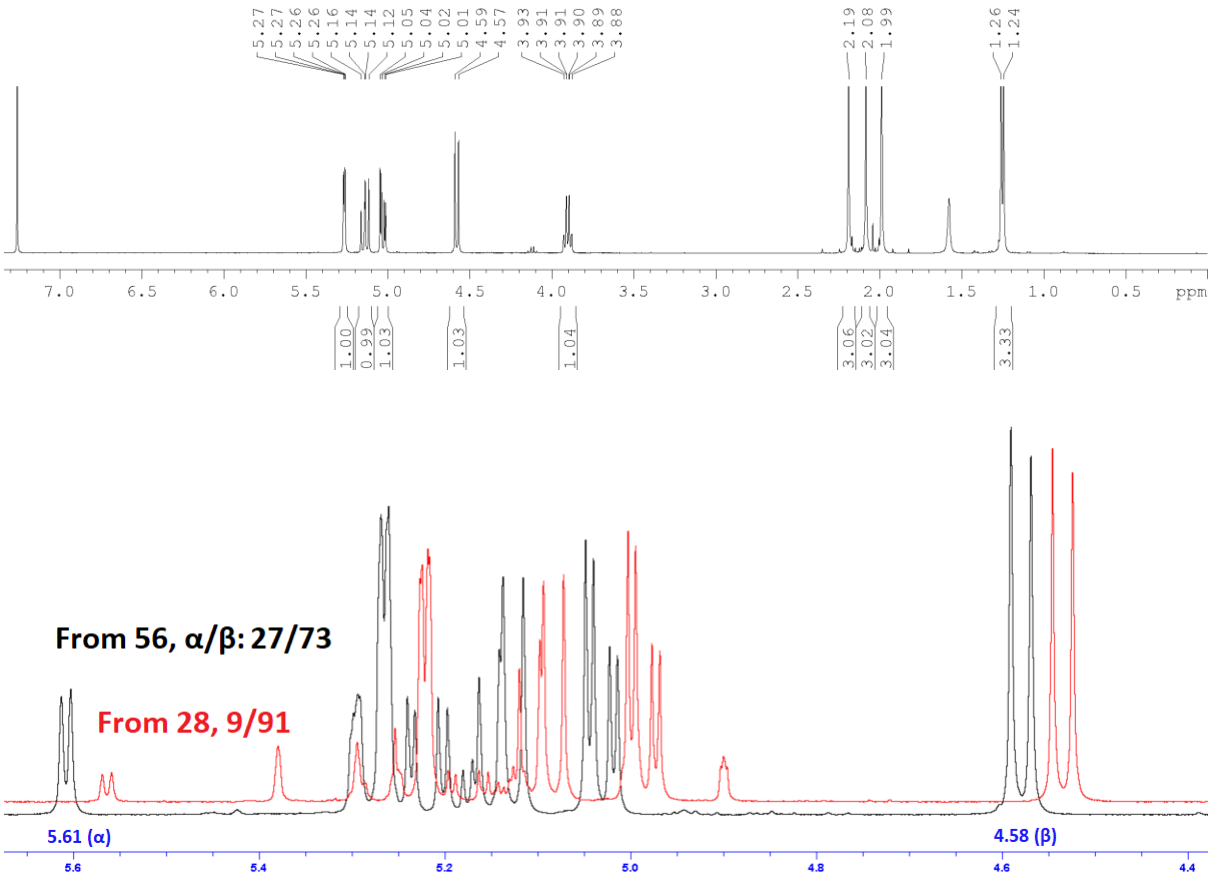
From **56**:



To a solution of **L-fucopyranosyl azide 56** (218 mg, 1.15 mmol, 1 eq) in pyridine (2 mL) was added acetic anhydride (2 mL). The reaction mixture was stirred at room temperature for 15 h until TLC showed completion, before diluted with DCM and washed with a 0.02 M HCl aqueous solution. The organic phase was dried over Na_2SO_4 and concentrated, affording a crude anomeric mixture (α/β ratio: 27/73), as observed by ^{1}H NMR (α/β anomeric protons: 5.61/4.58 ppm). A portion (103 mg) of the crude was purified by flash chromatography (nHex/EtOAc 8:2) affording product **29** (42 mg, 0.13 mmol, y = 40 %).

^{1}H NMR (400 MHz, $CDCl_3$):

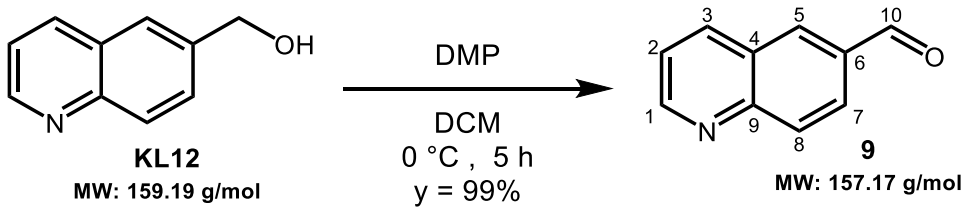
δ = 5.27 (dd, $J_{4\text{-}3}$ = 3.4 Hz, $J_{4\text{-}5}$ = 1.2 Hz, 1H, *H-4*), 5.14 (dd, $J_{2\text{-}1}$ = 8.6 Hz $J_{2\text{-}3}$ = 10.4 Hz, 1H, *H-2*), 5.03 (dd, $J_{3\text{-}2}$ = 10.4 Hz, $J_{3\text{-}4}$ = 3.4 Hz, 1H, *H-3*), 4.58 (d, $J_{1\text{-}2}$ = 8.6 Hz, 1H, *H-1*), 3.90 (dq, $J_{5\text{-}4}$ = 1.2 Hz, $J_{5\text{-}CH3}$ = 6.4 Hz, 1H, *H-5*), 2.19 (s, 3H, *OAc*), 2.08 (s, 3H, *OAc*), 1.99 (s, 3H, *OAc*), 1.25 (d, $J_{CH3\text{-}5}$ = 6.4 Hz, 3H, *CH_3*). In accordance with published data.[221]



General procedure for oxidation of fragments KL09 - KL12 following the procedure of George and co-workers:[223]

The **fragment** (1 eq) was dissolved in DCM (concentration: 0.1 M) and cooled to 0 °C. A DMP 15% w/w DCM solution (1.7 eq) was added to the solution and left to stir until TLC showed completion, while returning to room temperature. The solvents were evaporated and the resulting crude was purified by flash chromatography (nHex/EtOAc: 6/4) affording the aldehyde.

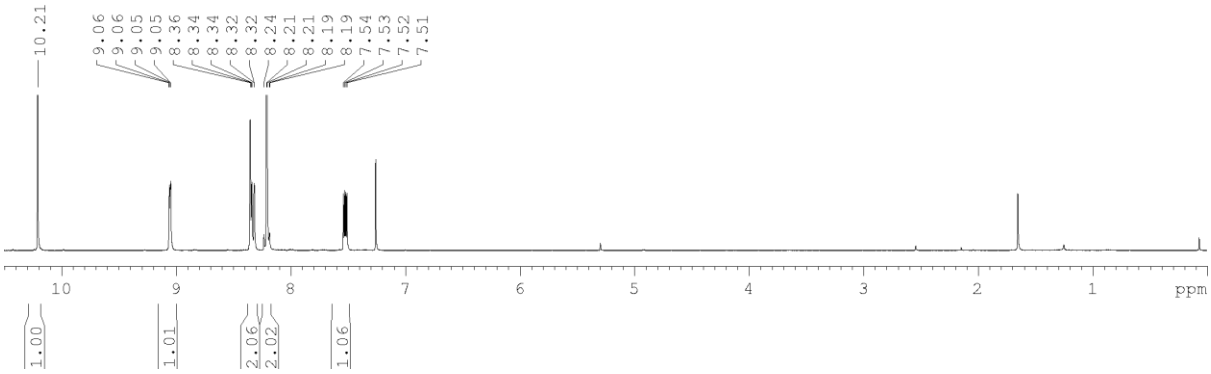
(9) Synthesis and characterization of **quinoline-6-carbaldehyde (9)**:



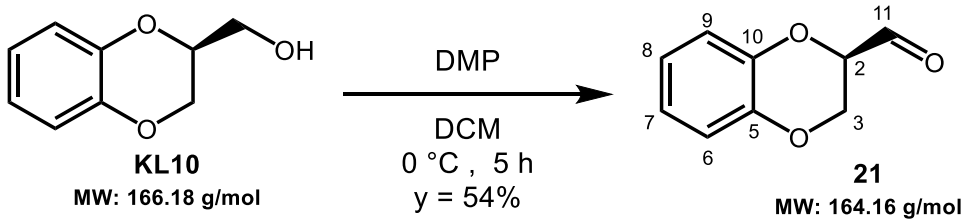
KL12: 6-quinolinylmethanol (0.28 mmol) was oxidized following the aforementioned procedure to afford **9** (0.28 mmol, y = 99 %). TLC R_f (nHex/EtOAc: 6/4): 0.24. MS (ESI) calculated for $C_{10}H_7NO$ $[M + H]^+$ *m/z*: 158.06; found: 158.00.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 10.21 (s, 1H, *H-10*), 9.05 (d, 1H, *H-1*), 8.34 (mult., 2H, *H-3* + *H-5*), 8.21 (mult., 2H, *H-7* + *H-8*), 7.52 (m, 1H, *H-2*). In accordance with published data.[224]



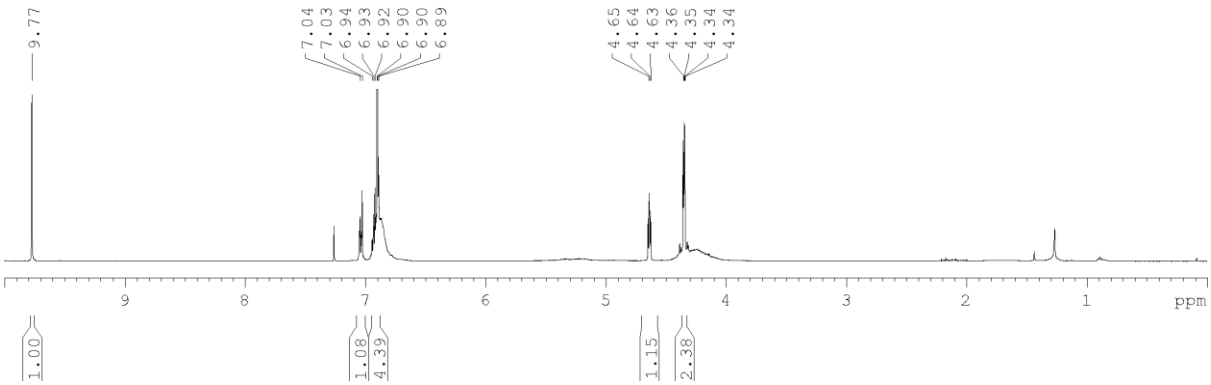
(21) Synthesis and characterization of **(R)-1,4-benzodioxane-2-carboxaldehyde (21)**:



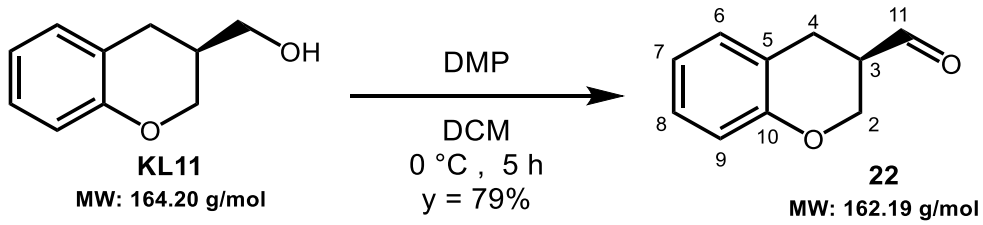
KL10: (S)-2-hydroxymethyl-1,4-benzodioxane (0.31 mmol) was oxidized following the aforementioned procedure to afford **21** (0.16 mmol, y = 54 %). TLC R_f (nHex/EtOAc: 6/4): 0.21.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 9.77 (s, 1H, *H-11*), 7.04 (m, 1H, *Ar*), 6.90 (mult., 3H, *Ar*), 4.64 (t, 1H), 4.35 (m, 2H). In accordance with published data.[195] The aldehyde hydrate is also visible in the spectrum.



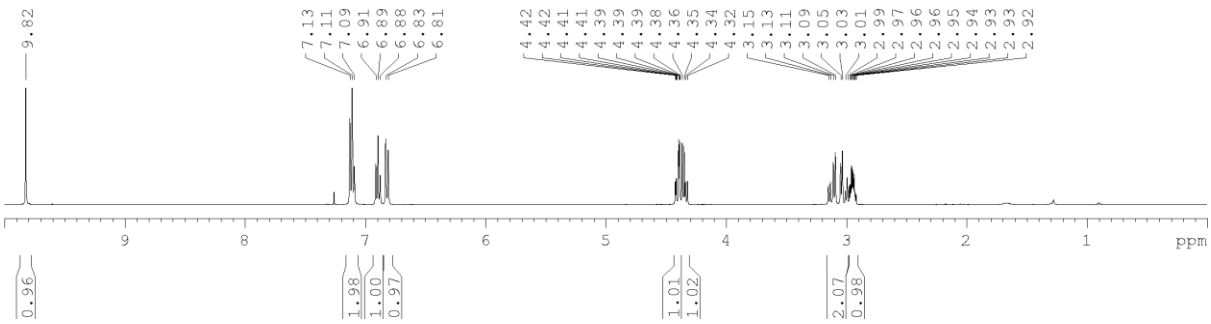
(22) Synthesis and characterization of **(R)-chromane-3-carbaldehyde (22)**:



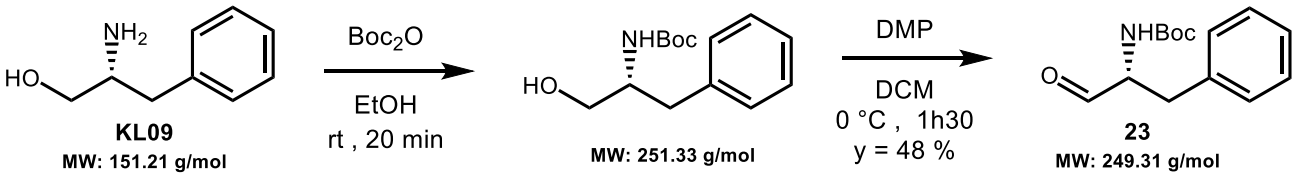
KL11: (-)-S-3-hydroxymethylchromane (0.30 mmol) was oxidized following the aforementioned procedure to afford **22** (0.23 mmol, y = 79 %). TLC R_f (nHex/EtOAc: 6/4): 0.56.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 9.82 (d, $J_{11\text{-}3}$ = 0.8 Hz, 1H, *H-11*), 7.11 (mult., 2H, *Ar*), 6.89 (t, 1H, *Ar*), 6.82 (d, J = 8.5 Hz, 1H, *Ar*), 4.40 (dd, J = 3.4 Hz, J' = 1.1 Hz, 1H), 4.34 (dd, J' = 1.1 Hz, J'' = 6.4 Hz, 1H), 3.07 (qd, 2H), 2.95 (m, $J_{3\text{-}11}$ = 0.8 Hz, 1H, *H-3*).



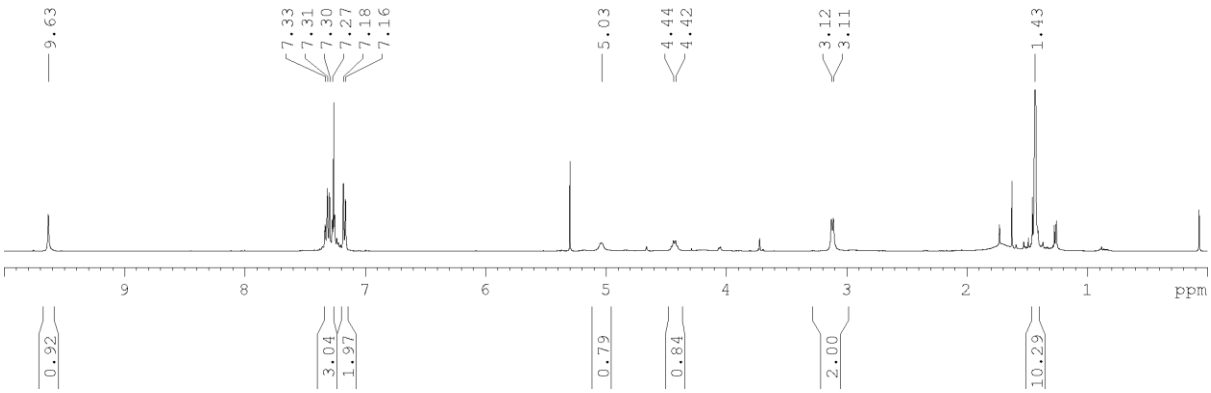
(23) Synthesis and characterization of ***tert*-butyl (R)-(1-oxo-3-phenylpropan-2-yl)carbamate (23)** following the procedures of Vilaivan, and Reggelin and co-workers:[225-226]



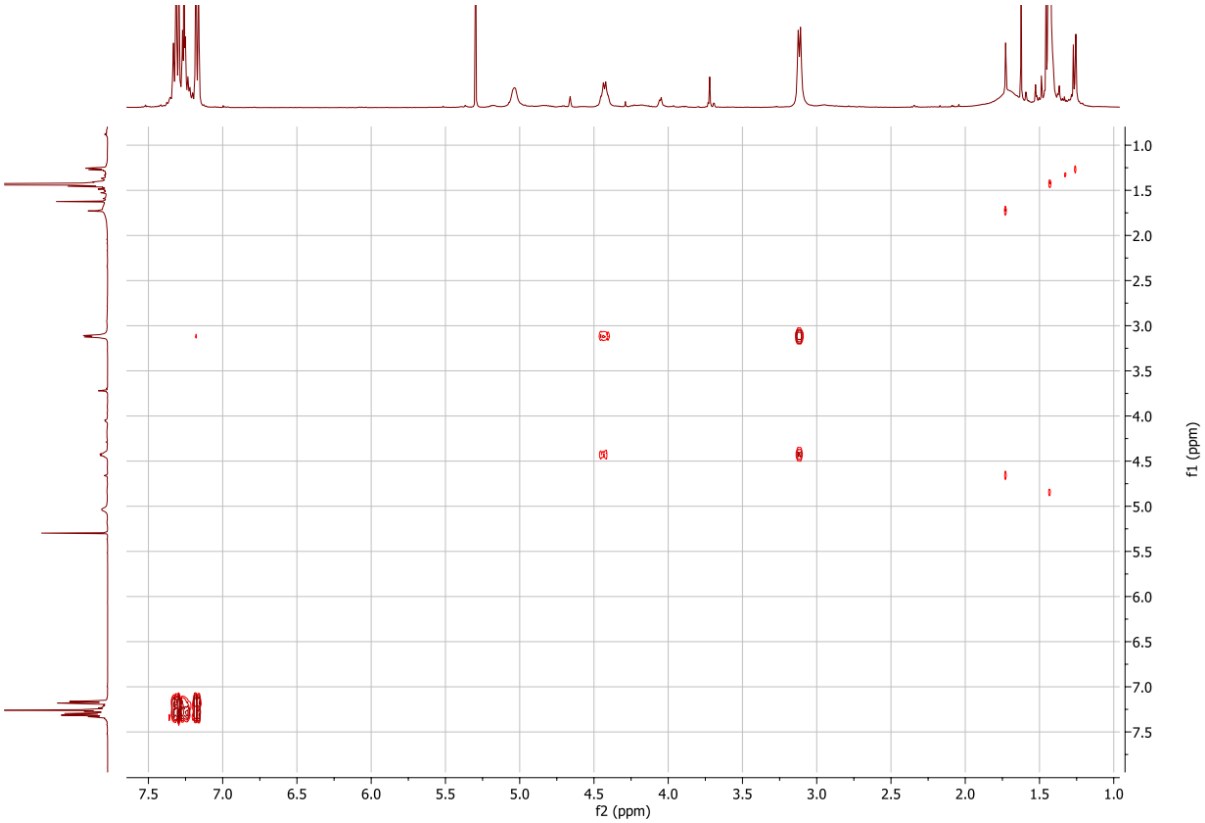
To a solution of **KL09: D-phenylalaninol** (101 mg, 0.67 mmol, 1 eq) in EtOH (2 mL) was added Boc_2O (192 µL, 0.84 mmol, 1.25 eq) under N_2 atmosphere. The reaction was stirred at room temperature for 20 min until TLC showed completion, before the solvents were concentrated. The crude (168 mg, quantitative) underwent oxidation following the aforementioned procedure to afford **23** (81 mg, 0.32 mmol, y = 48 %). TLC R_f intermediate (DCM/MeOH: 9/1): 0.67. TLC R_f product (nPent/EtOAc: 6/4): 0.50.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 9.63 (d, 1H, *O=CH*), 7.30 (mult., 3H, *CH Ar*), 7.17 (d, J = 7.4 2H, *CH Ar*), 5.03 (bs, 1H, *NH*), 4.43 (m, J = 1.1 Hz, J = 6.4 Hz, 1H, *N-CH*), 3.12 (d, J = 6.6 Hz, 2H, CH_2), 1.43 (bs, 9H, *tBu*).

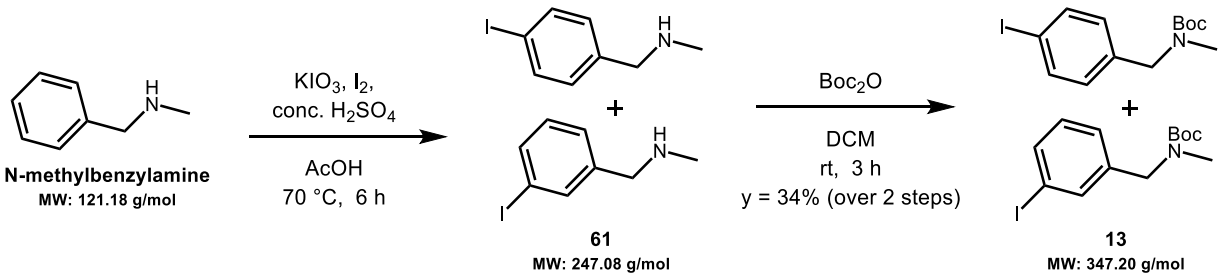


COSY:



(13) Synthesis and characterization of ***tert*-butyl (4-iodobenzyl)(methyl)carbamate (13)**:

Method 1: from N-methylbenzylamine following the procedure of Lei and co-workers (isomeric mixture):[227-228]

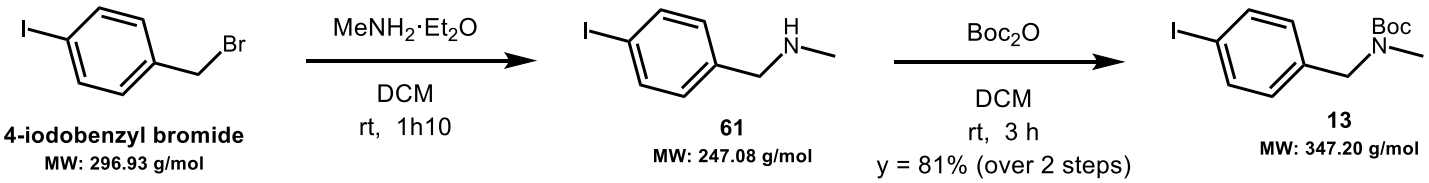


N-methylbenzylamine (107 mg, 0.83 mmol, 1 eq), I_2 (84 mg, 0.33 mmol, 0.4 eq) and KIO_3 (99 mg, 0.46 mmol, 0.5 eq) were dissolved in AcOH (750 µL), then stirred at room temperature for 30 min. Concentrated H_2SO_4 (100 µL, 1.88 mmol, 2.1 eq) was added and the reaction mixture was refluxed (70 °C) for 6 h until TLC showed completion, before returning to room temperature. The mixture was diluted with water and neutralized with a 5 M NaOH aqueous solution, then extracted with DCM. The organic phase was dried over Na_2SO_4 and concentrated. The crude **61** (97 mg) consisted of a mixture of *para*/*meta* regioisomers as seen by ^{1}H NMR (see details below for **13**), and was used for the next step without purification. TLC

R_f (DCM/MeOH: 8/2): 0.13. MS (ESI) calculated for $C_8H_{10}IN$ $[M + H]^+$ *m/z*: 247.99; found: 247.92.

The crude **N-methyl-(3/4-iodobenzyl)amine 61** (97 mg, max: 0.39 mmol, 1 eq) was dissolved in DCM (4 mL) and stirred at room temperature under N_2 atmosphere. Boc_2O (120 mg, 0.55 mmol, 1.4 eq) was added and the reaction mixture was stirred for 3 h until TLC showed completion, before the solvent was concentrated. The crude product was purified by flash chromatography (nHex/EtOAc: 9:1) affording product **13** (105 mg, 0.30 mmol, y = 34 % over 2 steps) as a *para*/*meta* 43:57 regioisomeric mixture (*para* signals at δ = 7.62, 6.96 ppm, *meta* at 7.57, 7.16, 7.04 ppm). TLC R_f (nHex/EtOAc: 8/2 and 9/1): 0.48 and 0.25.

Method 2: from 4-iodobenzyl bromide (pure para-iodide)



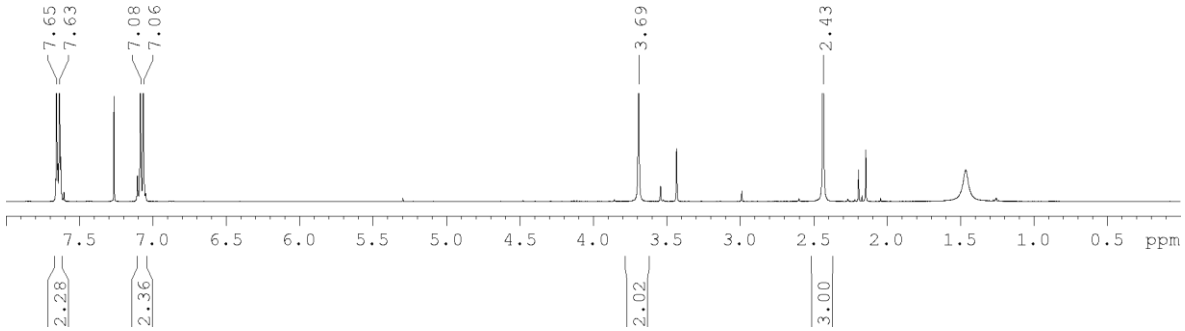
4-iodobenzyl bromide (2.0 g, 6.74 mmol, 1 eq) was dissolved in DCM under N_2 atmosphere. The solution was transferred to a dropping funnel and added to a stirring $MeNH_2$ 33% w/w Et_2O solution (12.5 mL, 100 mmol, 14.8 eq) over 30 min. The reaction was stirred at room temperature for an additional 40 min until TLC showed completion, before concentrating. The crude was dissolved in DCM and washed with a NaOH aqueous solution. The organic phase was dried over Na_2SO_4 and concentrated to crude **61** (1.56 g), used for the next step without purification. TLC R_f (nHex/EtOAc: 9/1): 0.60.

Crude **N-methyl-(4-iodobenzyl)amine 61** (2.01 g, max: 8.14 mmol, 1 eq) was dissolved in DCM (81 mL) and stirred at room temperature under N_2 atmosphere. A solution of Boc_2O (2.47 mg, 11.32 mmol, 1.4 eq) in DCM was added and the reaction mixture was stirred for 3h until TLC showed completion, before the solvent was concentrated. The crude product was purified by automatic chromatography (Biotage Sfär 100: nHex/EtOAc 10% isocratic) affording product **13** (2.293 g, 6.60 mmol, y = 81 % over 2 steps). TLC R_f (nHex/EtOAc: 9/1): 0.25.

Crude N-methyl-(4-iodobenzyl)amine **61**:

^1H NMR (400 MHz, $CDCl_3$):

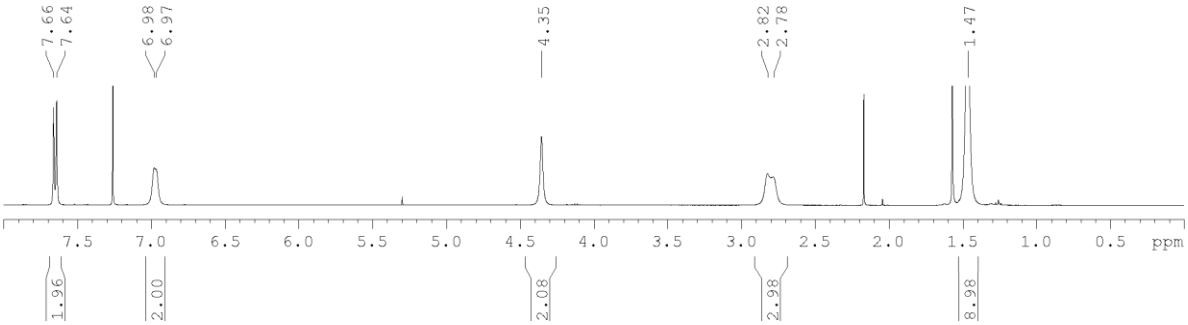
δ = 7.64 (d, J = 8.3 Hz, 2H, <u>*CH*</u>-*C-I*), 7.07 (d, J = 8.2 Hz, 2H, *CH Ar*), 3.69 (s, 2H, CH_2), 2.43 (s, 3H, CH_3).



tert-butyl (4-iodobenzyl)-N-(methyl)carbamate **13**:

^{1}H NMR (400 MHz, $CDCl_3$):

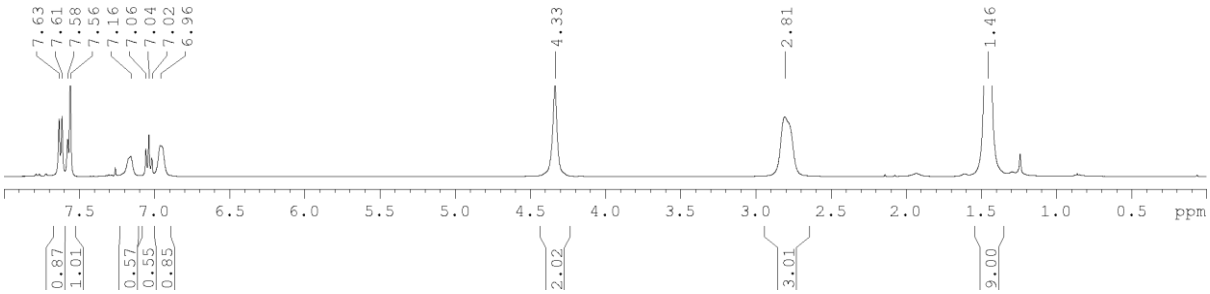
δ = 7.65 (d, J = 8.2 Hz, 2H, <u>*CH*</u>-*C-I*), 6.97 (bd, 2H, *CH Ar*), 4.35 (s, 2H, CH_2), 2.80 (bd, 3H, CH_3), 1.47 (bs, 9H, *tBu*). In accordance with published data.[229]



Isomeric mixture of Tert-butyl (4-iodobenzyl)(methyl)carbamate **13**: *para*/*meta* ratio (43:57)

^{1}H NMR (400 MHz, $CDCl_3$):

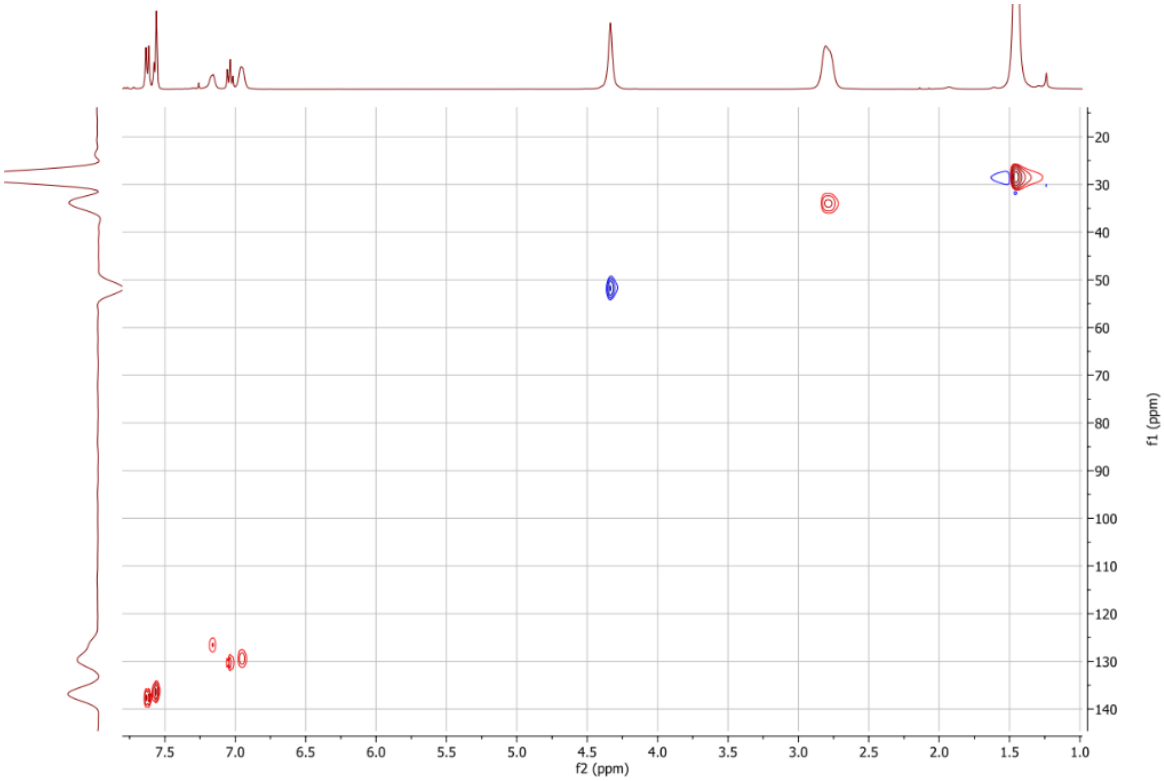
δ = 7.62 (d, 2H, *para*), 7.57 (d, 2H, *meta*), 7.16 (bs, 1H, *meta*), 7.04 (t, 1H, *meta*), 6.96 (bd, 2H, *para*), 4.33 (s, 2H, CH_2), 2.81 (bd, 3H, CH_3), 1.46 (bs, 9H, *tBu*).



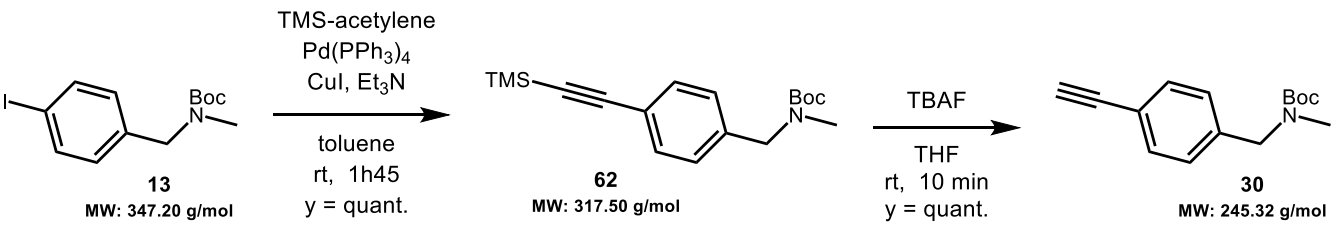
^{13}C NMR shifts extrapolated from the HSQC experiment:

para isomer: δ = 137.6 (*CH-C-I Ar*), 129.4 (*CH Ar*), 51.7 (CH_2), 33.9 (CH_3), 28.4 (*tBu*).

HSQC:



(30) Synthesis and characterization of ***tert*-butyl methyl(4-((ethynyl)benzyl)carbamate (30)** following the procedure of Decréau and co-workers:[230]



The pure *para*-iodinated isomer of **13** (synthesized by Method 2) ***tert*-butyl (4-iodobenzyl)(methyl)carbamate** (197 mg, 0.57 mmol, 1 eq), $Pd(Ph_3)_4$ (24 mg, 0.02 mmol, 0.05 eq) and CuI (16 mg, 0.09 mmol, 0.2 eq) were dissolved in toluene (1.5 mL), under Ar atmosphere. TMS-acetylene (100 µL, 0.70 mmol, 1.2 eq) and Et_3N (100 µL, 0.72 mmol, 1.3 eq) were added and the reaction mixture was stirred at room temperature for 1h45 until TLC showed completion, before being concentrated. The crude was purified by automatic chromatography (Biotage Sfär 25: nHex/EtOAc gradient 0 % to 25 %) affording intermediate **62** (184 mg, quantitative yield). TLC R_f (nHex/EtOAc: 8/2): 0.70. MS (ESI) calculated for $C_{18}H_{27}NO_2Si$ $[M + Na]^+$ *m/z*: 340.17; found: 340.09.

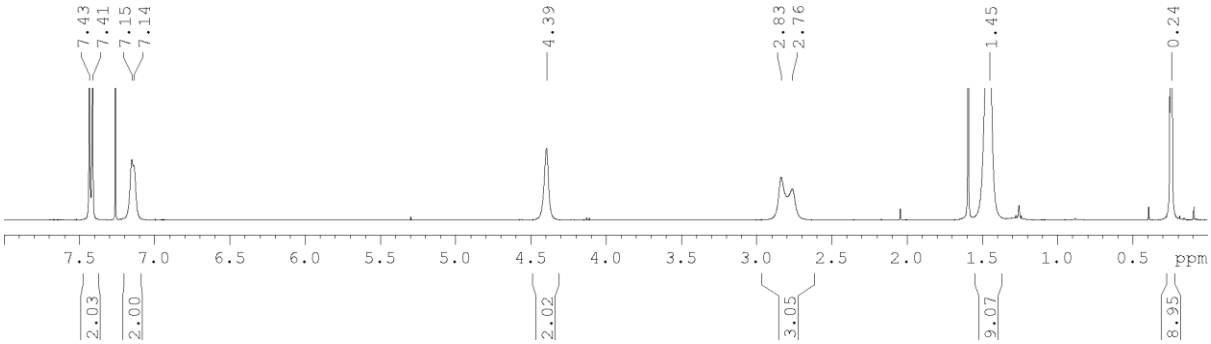
To ***tert*-butyl methyl(4-((trimethylsilyl)ethynyl)benzyl)carbamate 62** (70 mg, 0.22 mmol, 1 eq) in THF (500 µL) was added a 1 M solution of TBAF in THF (250 µL, 0.25 mmol, 1.1 eq). The

brown reaction mixture was stirred at room temperature for 10 min until TLC showed completion, before being concentrated. The crude was dissolved in DCM and washed with a 1 M aqueous HCl solution. The organic phase was dried over Na_2SO_4 and concentrated to crude ***tert*-butyl methyl(4-((ethynyl)benzyl)carbamate 30** (55 mg, quantitative yield), used for the next step without purification. TLC R_f (nHex/EtOAc: 8/2): 0.56. MS (ESI) calculated for $C_{15}H_{19}NO_2$ $[M + Na]^+$ *m/z*: 268.13; found: 268.11.

Crude *tert*-butyl methyl(4-((trimethylsilyl)ethynyl)benzyl)carbamate **62**:

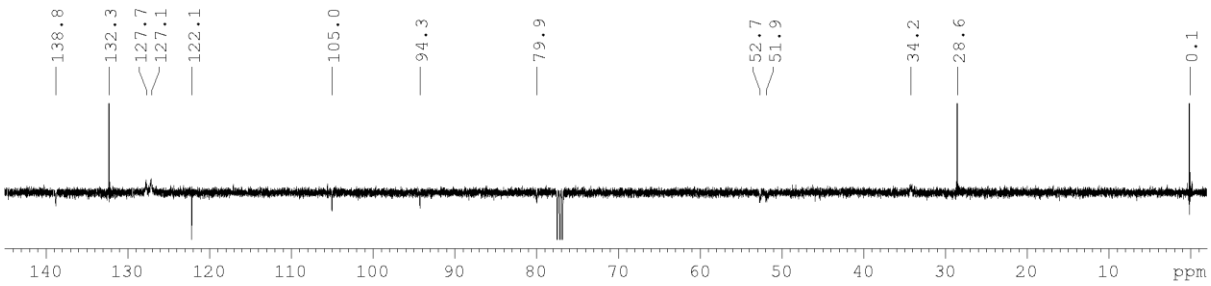
^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.42 (d, J = 8.3 Hz, 2H, ≡*C-C-CH Ar*), 7.14 (d, J = 6.9 Hz, 2H, *CH Ar*), 4.39 (bs, 2H, CH_2), 2.83 - 2.76 (bd, 3H, CH_3), 1.47 (bs, 9H, *tBu*), 0.24 (s, 9H, Si-CH_3).

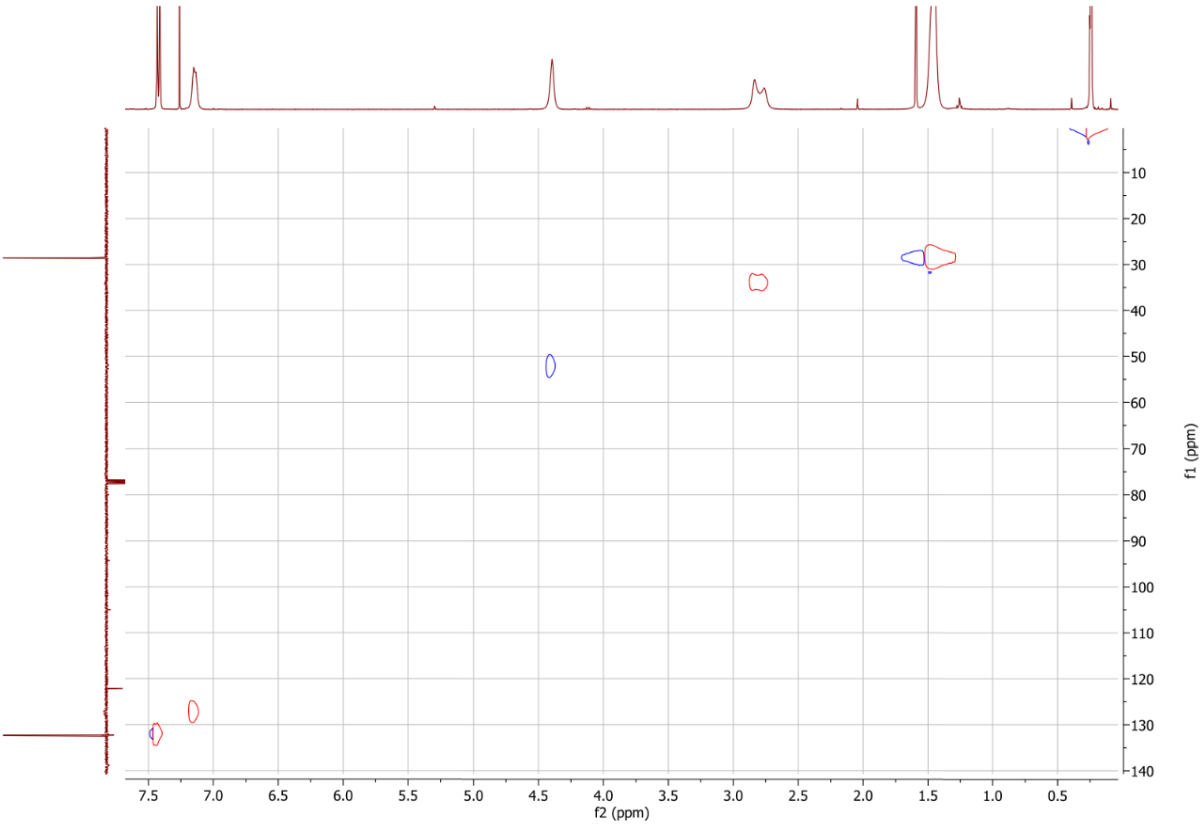


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

δ = 138.8 (CH_2-*C Ar*), 132.3 (≡*C-C-CH Ar*), 127.7 - 127.1 (*CH Ar*), 122.1 (≡*C-C Ar*), 105.0 (≡*C-C Ar*), 94.3 (≡*C-Si*), 52.7 - 51.9 (CH_2), 34.2 (CH_3), 28.6 (*tBu*), 0.1 (*Si-*CH_3).



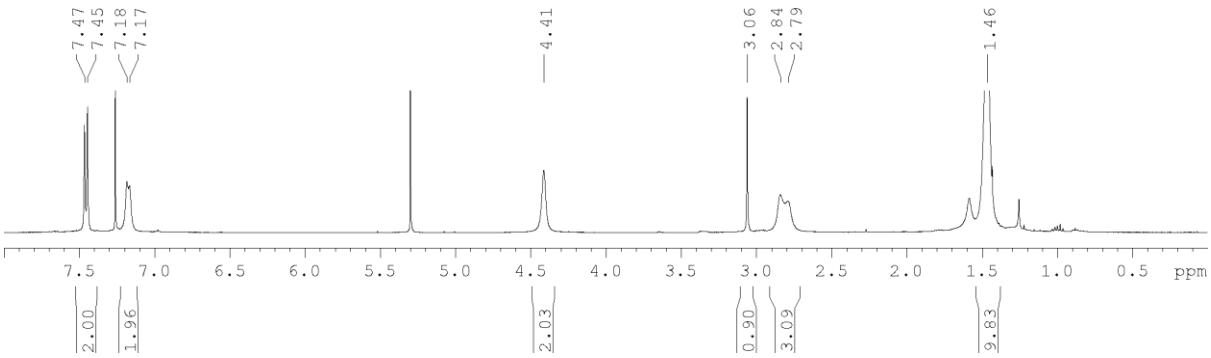
HSQC:



tert-butyl methyl(4-((ethynyl)benzyl)carbamate **30**:

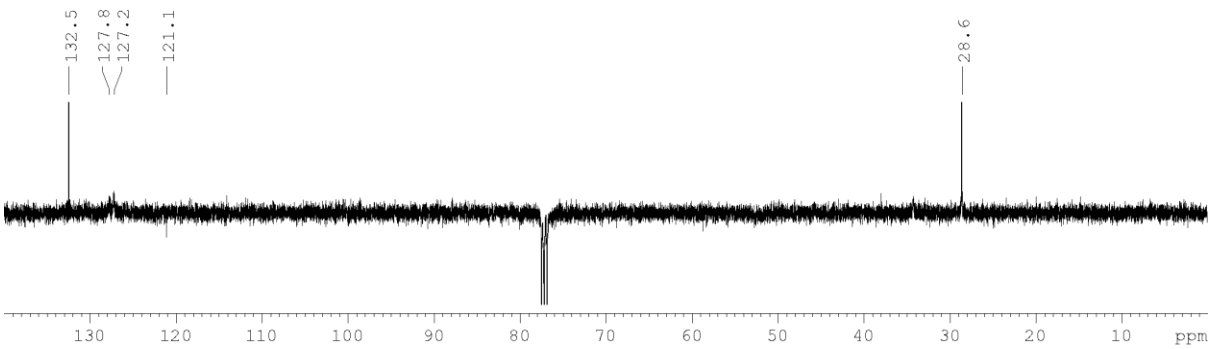
^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.46 (d, J = 8.1 Hz, 2H, ≡*C-C-CH Ar*), 7.17 (d, J = 6.9 Hz, 2H, *CH Ar*), 4.41 (bs, 2H, CH_2), 3.06 (s, 1H, ≡*CH*), 2.84 - 2.79 (bd, 3H, CH_3), 1.46 (bs, 9H, *tBu*).

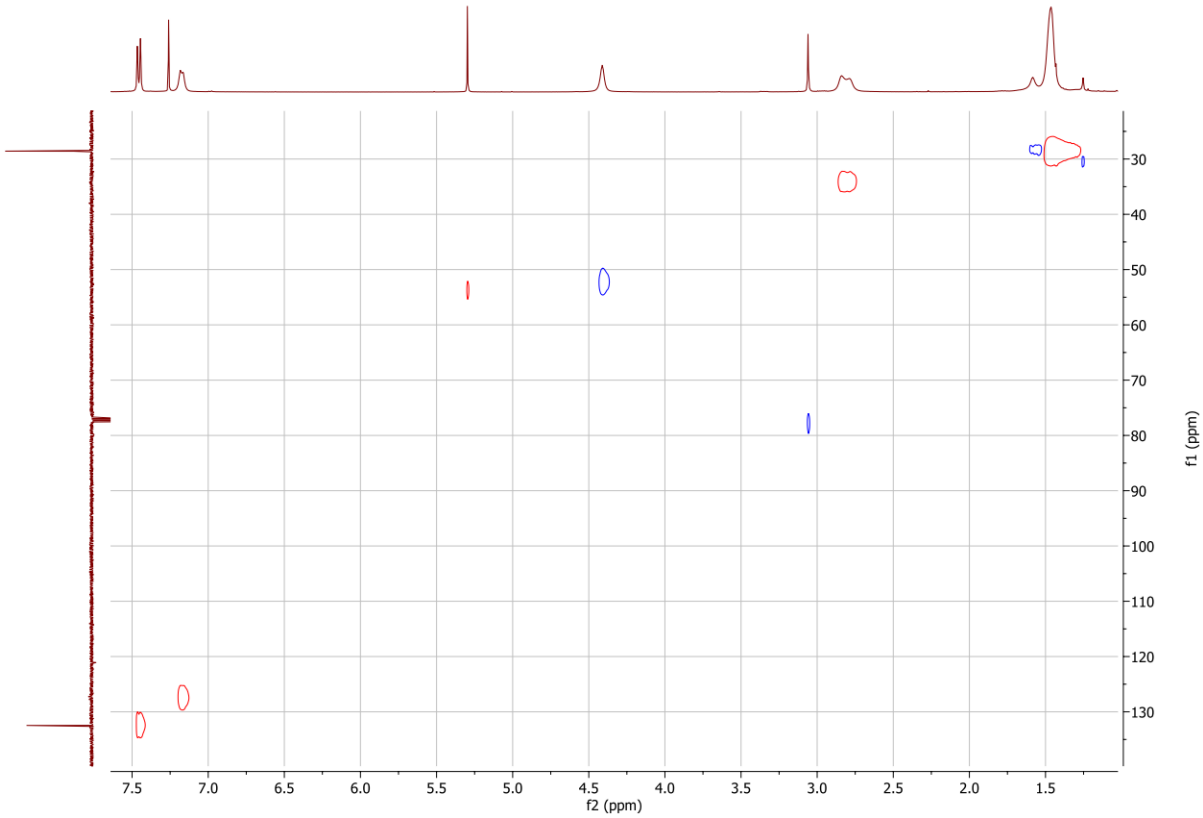


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

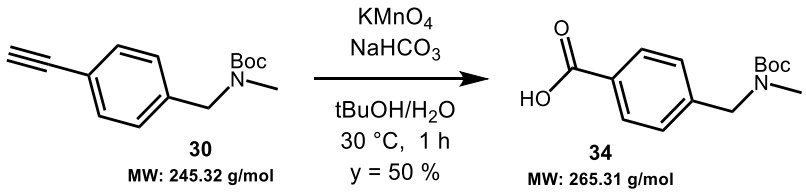
δ = 132.3 (≡*C-C-CH Ar*), 127.8 - 127.2 (*CH Ar*), 121.1 (≡*C-C Ar*), 77.7 (≡*CH*), 52.2 (CH_2), 33.8 (CH_3), 28.6 (*tBu*).



HSQC:



(34) Synthesis and characterization of **4-(((tert-butoxycarbonyl)(methyl)amino)methyl) benzoic acid (34)**:

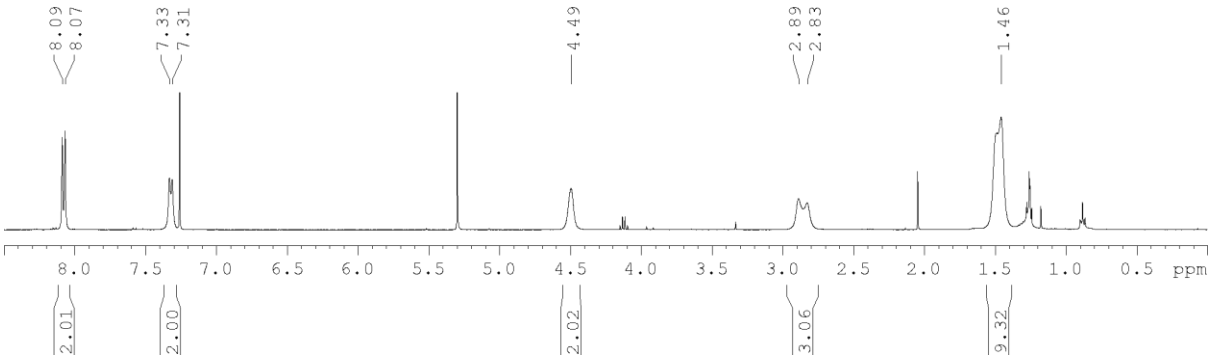


***tert*-butyl (4-iodobenzyl)(methyl)carbamate 30** (82 mg, 0.33 mmol, 1 eq) was dissolved in tBuOH (11 mL). Separately, $KMnO_4$ (158 mg, 1.00 mmol, 3 eq) was dissolved in water (11 mL). Separately, $NaHCO_3$ (87 mg, 1.04 mmol, 3 eq) was dissolved in water (11 mL). The $NaHCO_3$ solution, followed by the $KMnO_4$ solution were added to the dissolved alkyne and stirred at

30 °C for 1 h until TLC showed completion. The reaction mixture was quenched by stirring with MeOH, then was filtered and concentrated. The crude was redissolved in DCM and acidified water (to pH 2, using 1M HCl). The organic phase was dried over Na_2SO_4 and concentrated. The crude was purified by automatic chromatography (Biotage Sfär 10: nHex/EtOAc gradient 0 % to 60 %; 0.01% formic acid) affording **34** (43 mg, 0.16 mmol, y = 50 %). TLC R_f (nHex/EtOAc: 7/3): 0.73. MS (ESI) calculated for $C_{14}H_{19}NO_4$ $[M + Na]^+$ *m/z*: 288.12; found: 288.14.

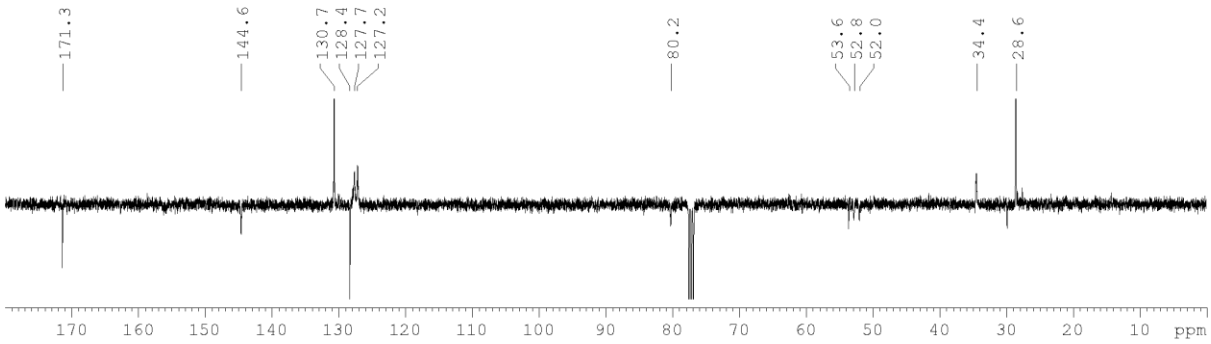
^{1}H NMR (400 MHz, $CDCl_3$):

δ = 8.08 (d, J = 8.2 Hz, 2H, O=*C-C-<u>CH</u> Ar*), 7.32 (d, J = 7.7 Hz, 2H, *CH Ar*), 4.49 (bs, 2H, *CH_2*), 2.89 - 2.83 (bd, 3H, *CH_3*), 1.46 (bs, 9H, *tBu*).

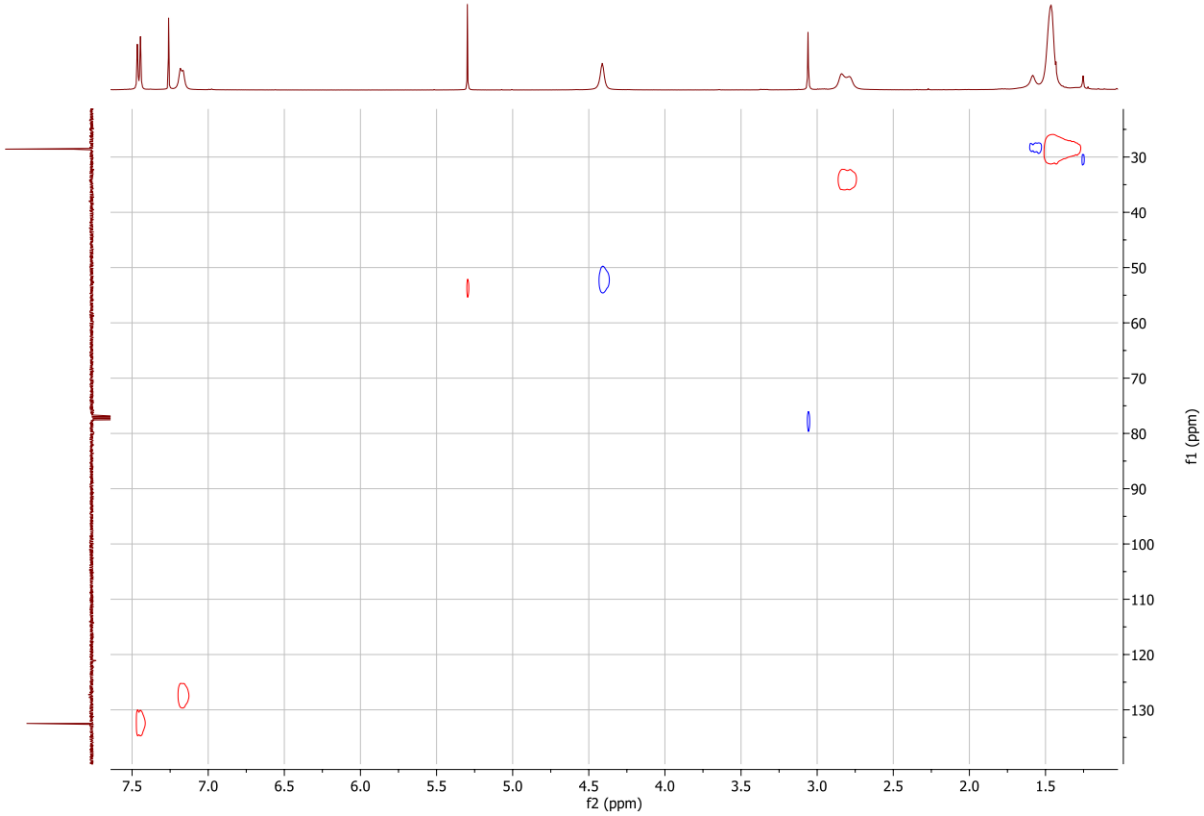


^{13}C NMR (400 MHz, $CDCl_3$):

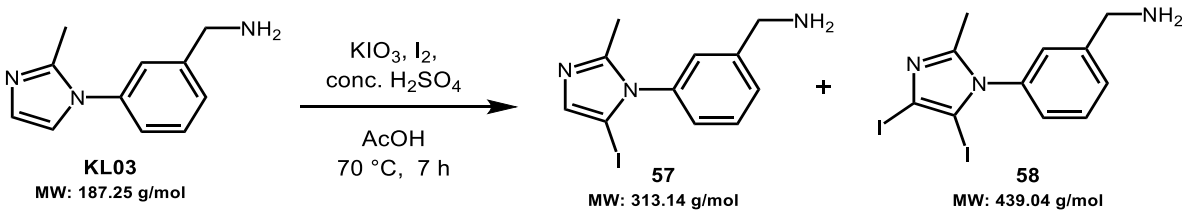
δ = 171.3 (*COOH*), 144.6 (*CH_2-<u>C</u> Ar*), 130.7 (O=*C-C-<u>CH</u> Ar*), 128.4 (O=*C-<u>C</u> Ar*), 127.7 - 127.2 (*CH Ar*), 80.2 (*C tBu*), 53.6 - 52.0 (*CH_2*), 34.4 (*CH_3*), 28.6 (*CH_3 tBu*).



HSQC:



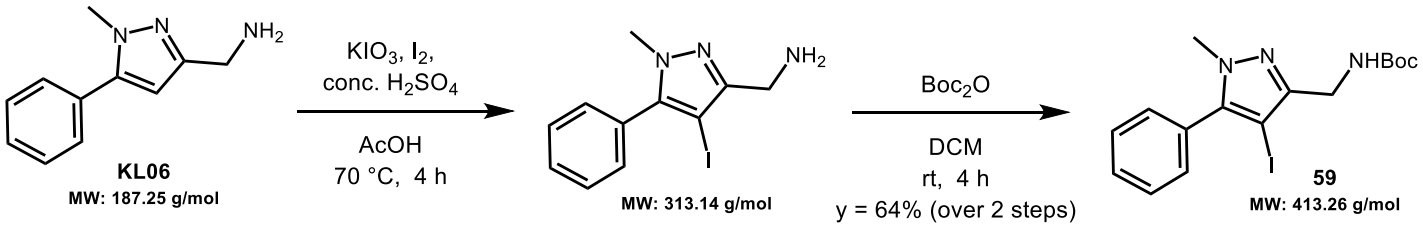
(57 + 58) Synthesis and characterization of **(3-(5-iodo-2-methyl-1H-imidazol-1-yl)phenyl) methanamine (57)** and **(3-(4,5-diiodo-2-methyl-1H-imidazol-1-yl)phenyl) methanamine (58)**:



(3-(2-methyl-1H-imidazol-1-yl)phenyl) methanamine - KL03 (30 mg, 0.16 mmol, 1 eq), I_2 (17mg, 0.07 mmol, 0.4 eq) and KIO_3 (10 mg, 0.05 mmol, 0.3 eq) were dissolved in AcOH (145 µL), then stirred at room temperature for 30 min. Concentrated H_2SO_4 (20 µL, 0.38 mmol, 2.3 eq) was added and the reaction mixture was refluxed (70 °C) for 7 h until TLC showed no further advancement: the starting material and two products were observed. The mixture returned to room temperature and was diluted with water, neutralized with a 5 M NaOH aqueous solution, and extracted with DCM. The organic phase was dried over Na_2SO_4 and concentrated (m = 26 mg). The crude intermediate was analyzed by MS: it revealed three species: the starting material and the mono and di-iodinated species, conceivably iodinated

in the positions described by Dubost and co-workers.[231] TLC R_f (DCM/MeOH: 8/2): 0.00, 0.22 and 0.36 for **KL03**, **57** and **58**. MS (ESI) calculated for **KL03** $C_{11}H_{13}N_3$ $[M + H]^+$ *m/z*: 188.11; found: 188.18. For **57** $C_{11}H_{12}IN_3$ $[M + H]^+$ *m/z*: 314.01; found: 314.23. For **58** $C_{11}H_{11}I_2N_3$ $[M + H]^+$ *m/z*: 439.91; found: 440.19.

(59) Synthesis and characterization of ***tert*-butyl ((4-iodo-1-methyl-5-phenyl-1H-pyrazol-3-yl)methyl)carbamate (59)**:



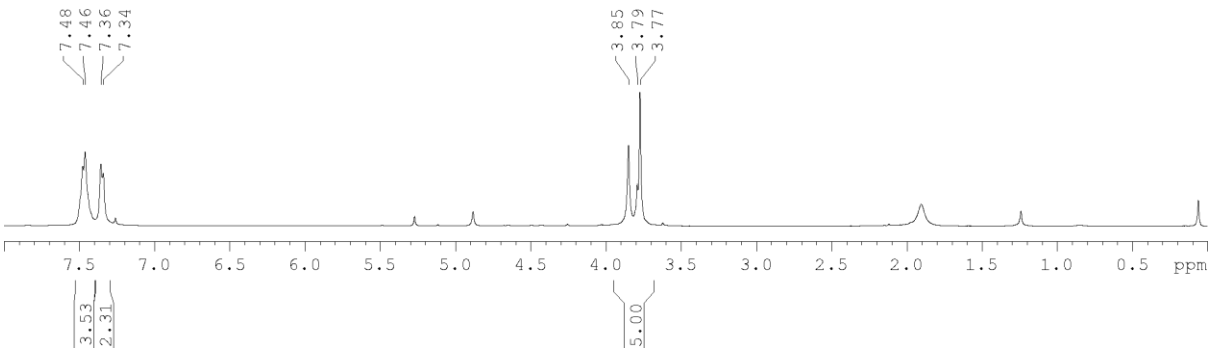
(1-methyl-5-phenyl-1H-pyrazol-3-yl)methanamine - KL06 (30 mg, 0.16 mmol, 1 eq), I_2 (16 mg, 0.06 mmol, 0.4 eq) and KIO_3 (10 mg, 0.05 mmol, 0.3 eq) were dissolved in AcOH (145 µL), then stirred at room temperature for 30 min. Concentrated H_2SO_4 (20 µL, 0.38 mmol, 2.3 eq) was added and the reaction mixture was refluxed (70 °C) for 4 h until TLC showed completion, before returning to room temperature. The mixture was diluted with water and neutralized with a 5 M NaOH aqueous solution, then extracted with DCM. The organic phase was dried over Na_2SO_4 and concentrated. The crude intermediate was not iodinated on the benzene ring, but on the pyrazole: **(4-iodo-1-methyl-5-phenyl-1H-pyrazol-3-yl) methanamine** (49 mg), was formed as seen by ^{1}H NMR (absent heterocyclic proton at δ = 6.5 ppm), and was used for the next step without purification. TLC R_f (DCM/MeOH: 8/2): 0.50. MS (ESI) calculated for $C_{11}H_{12}IN_3$ $[M + Na]^+$ *m/z*: 336.00; found: 355.98.

Crude **(4-iodo-1-methyl-5-phenyl-1H-pyrazol-3-yl) methanamine** (49 mg, max: 0.16 mmol, 1 eq) was dissolved in DCM (1.6 mL) and stirred at room temperature under N_2 atmosphere. Boc_2O (47 µL, 0.21 mmol, 1.3 eq) was added and the reaction mixture stirred for 4h until TLC showed completion, before the solvent was concentrated. The crude product was purified by flash chromatography (nHex/EtOAc: 7:3) affording product **59** (42 mg, 0.10 mmol, y = 64 % over 2 steps). TLC R_f (nHex/EtOAc: 7/3): 0.27.

Crude **(4-iodo-1-methyl-5-phenyl-1H-pyrazol-3-yl) methanamine**:

^{1}H NMR (400 MHz, $CDCl_3$):

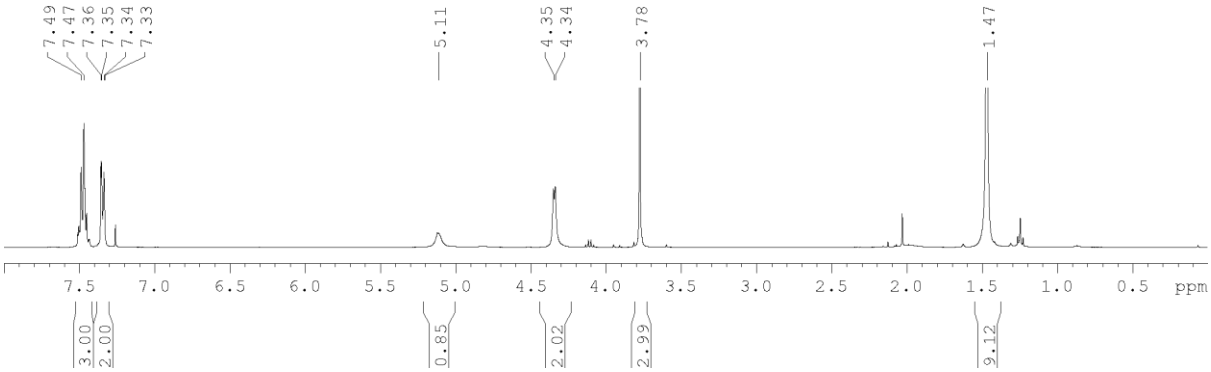
δ = 7.47 (mult., 3H, *CH Ar*), 7.3 (mult., 2H, *CH Ar*), 3.85 - 3.77 (mult. 5H, CH_2 + CH_3).



tert-butyl ((4-iodo-1-methyl-5-phenyl-1H-pyrazol-3-yl)methyl)carbamate **59**:

^{1}H NMR (400 MHz, $CDCl_3$):

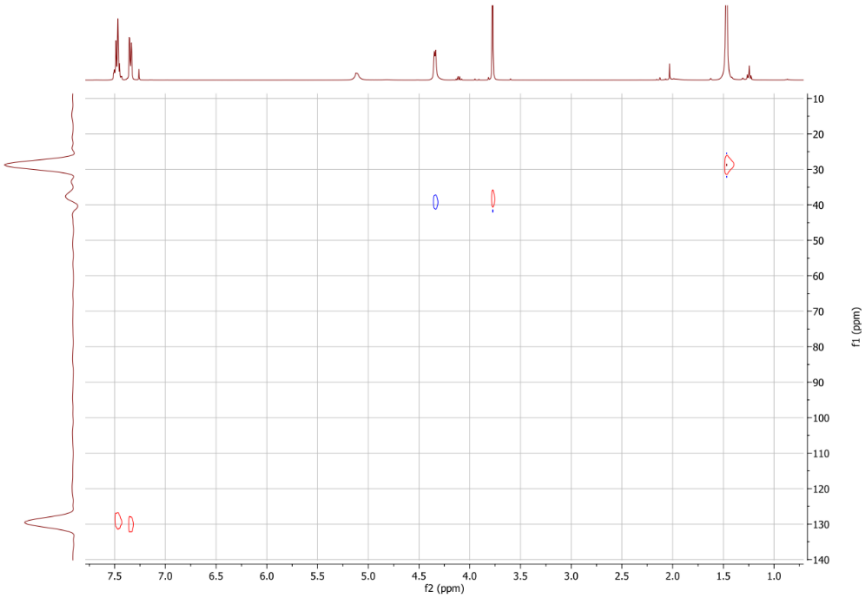
δ = 7.48 (mult., 3H, *CH Ar*), 7.34 (mult., 2H, *CH Ar*), 5.11 (*NH*), 4.35 (s, 2H, CH_2), 3.78 (s, 3H, CH_3), 1.47 (bs, 9H, *tBu*).



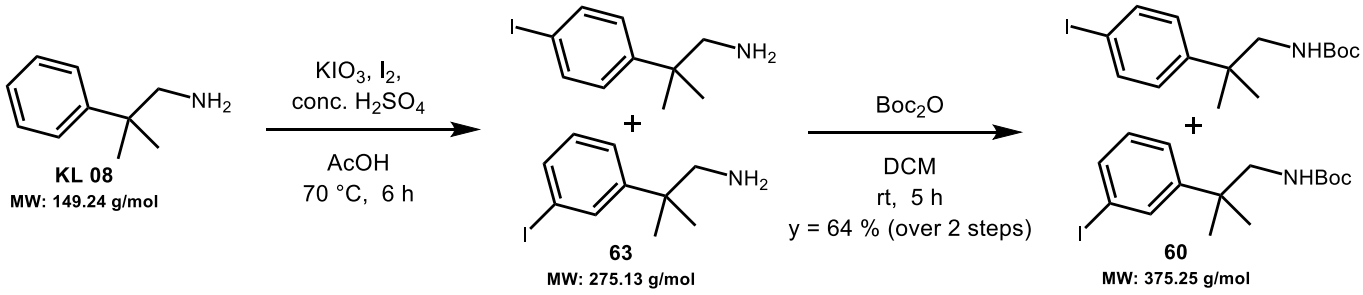
^{13}C NMR shifts extrapolated from the HSQC experiment:

δ = 130.1, 129.1 (*CH Ar*), 39.2 (CH_2), 38.3 (CH_3), 28.7 (*tBu*).

HSQC:



(60) Synthesis and characterization of ***tert*-butyl (2-(4-iodophenyl)-2-methylpropyl)carbamate (60)**:



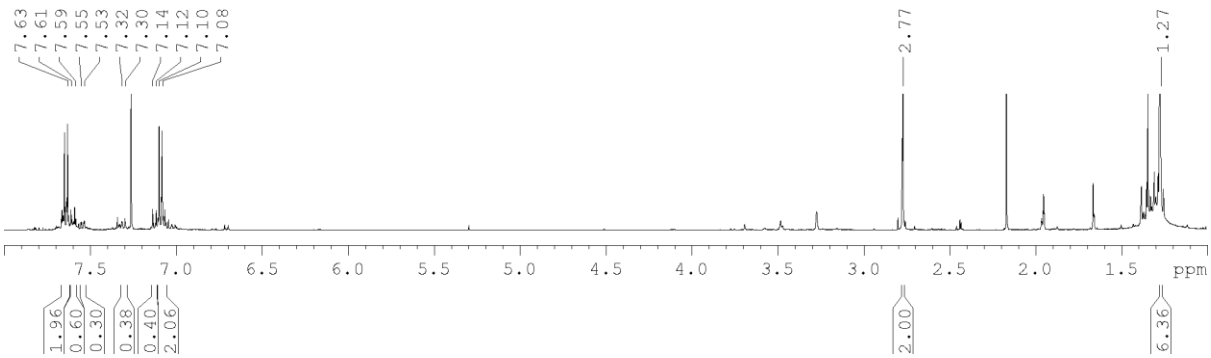
2-methyl-2-phenylpropan-1-amine - KL08 (97 mg, 0.65 mmol, 1 eq), I_2 (71 mg, 0.28 mmol, 0.4 eq) and KIO_3 (71 mg, 0.33 mmol, 0.5 eq) were dissolved in AcOH (610 µL), then stirred at room temperature for 30 min. Concentrated H_2SO_4 (80 µL, 1.47 mmol, 2.3 eq) was added and the reaction mixture was refluxed (70 °C) for 6 h until TLC showed completion, before returning to room temperature. The mixture was diluted with water and neutralized with a 5 M NaOH aqueous solution, then extracted with DCM. The organic phase was dried over Na_2SO_4 and concentrated. The crude **63** (177 mg) consisted of a mixture of regioisomers, as seen in the aromatic section of the ^{1}H NMR (difficult to approximate the ratio, see below for **60**), and was used for the next step without purification. TLC R_f (DCM/MeOH: 8/2): 0.65.

Crude **2-(4-iodophenyl)-2-methylpropan-1-amine 63** (177 mg, max: 0.643 mmol, 1 eq) was dissolved in DCM (6.4 mL) and stirred at room temperature under N_2 atmosphere. Boc_2O (183 mg, 0.84 mmol, 1.3 eq) was added and the reaction mixture stirred for 5 h until TLC showed completion, before the solvent was concentrated. The crude product was purified by flash chromatography (nHex/EtOAc: 9:1) affording product **60** (156 mg, 0.42 mmol, y = 64 % over 2 steps) as a regioisomeric mixture (*para*/*meta* 70:30), as seen by ^{1}H NMR (*para* signals at δ = 7.65, 7.10 ppm, *meta* at 7.65, 7.56, 7.33, 7.06 ppm). TLC R_f (nHex/EtOAc: 9/1): 0.30. MS (ESI) calculated for $C_{15}H_{22}INO_2$ $[M + Na]^+$ *m/z*: 398.06; found: 398.09. MS (ESI) calculated for $C_{15}H_{22}INO_2$ $[M + Na]^+$ *m/z*: 376.08; found: 376.03.

Crude 2-(4-iodophenyl)-2-methylpropan-1-amine **63**:

^{1}H NMR (400 MHz, $CDCl_3$):

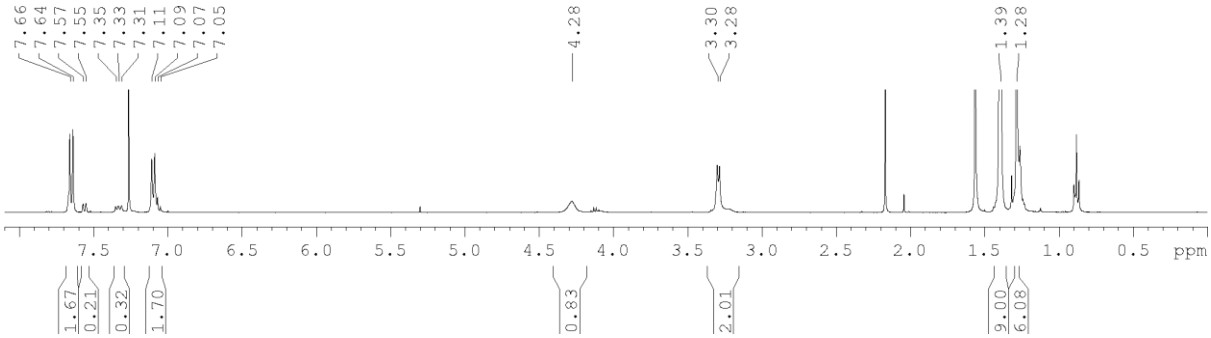
δ = 7.62 (d, 2H, *CH-C-I Ar*), 7.07 (d, 2H, *CH Ar*), 2.77 (s, 2H, CH_2), 1.27 (s, 6H, CH_3).



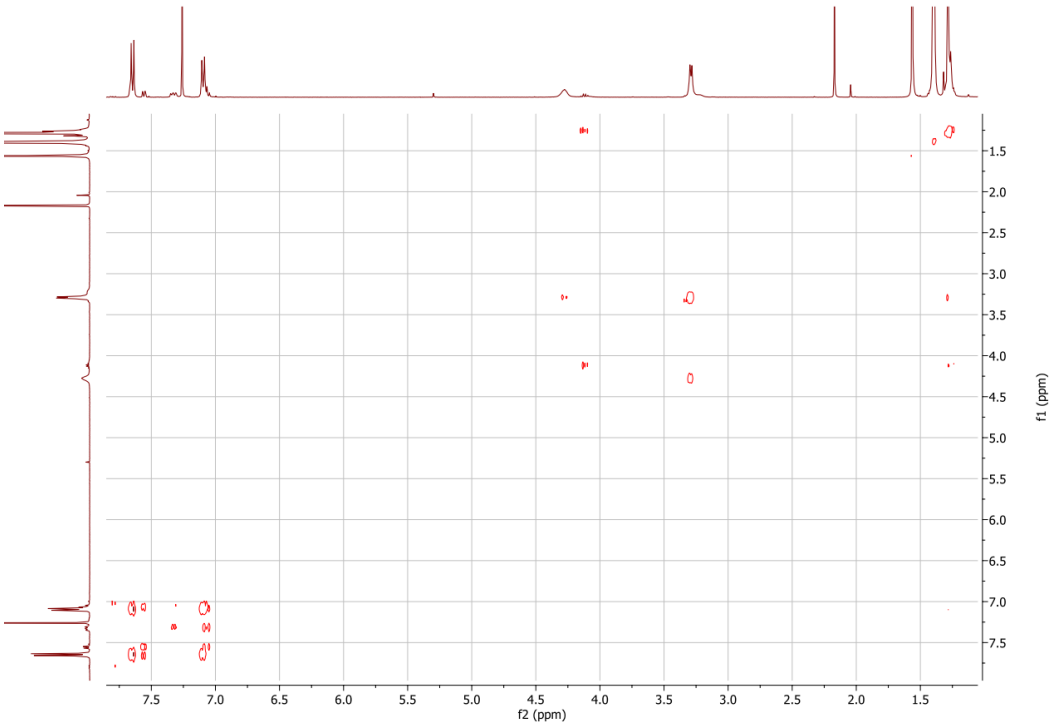
tert-butyl (2-(4-iodophenyl)-2-methylpropyl)carbamate **60**: *para*/*meta* ratio (70:30)

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.65 (d, J = 8.5 Hz, 2H, *para* <u>*CH*</u>*-C-I Ar*), 7.65 (m, 1H, *meta CH Ar*), 7.56 (d, J = 7.8 Hz, 1H, *meta CH Ar*), 7.33 (m, 1H, *meta CH Ar*), 7.10 (d, J = 8.4 Hz, 2H, *para CH Ar*), 7.06 (m, 1H, meta *CH Ar*), 4.28 (bs, 1H, *NH*), 3.29 (d, J = 6.3 Hz, 2H, CH_2), 1.39 (bs, 9H, *tBu*), 1.28 (s, 6H, CH_3).



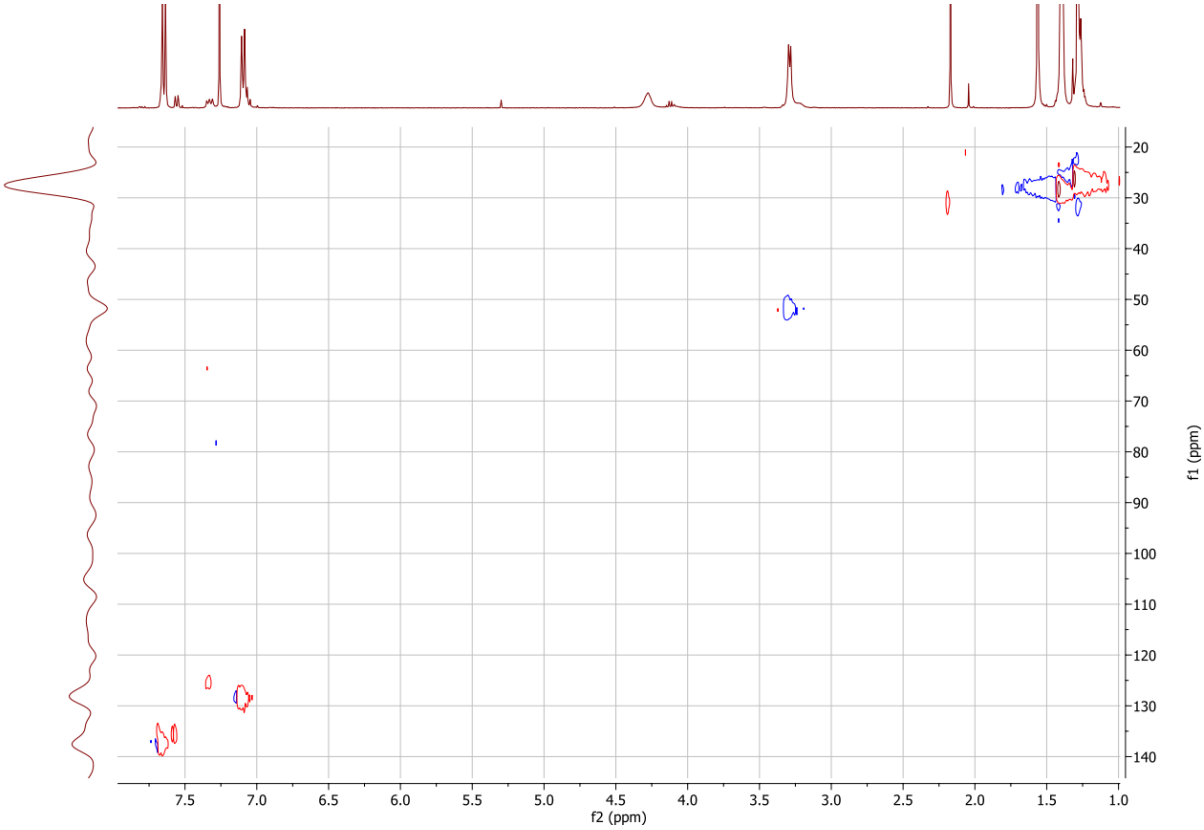
COSY:



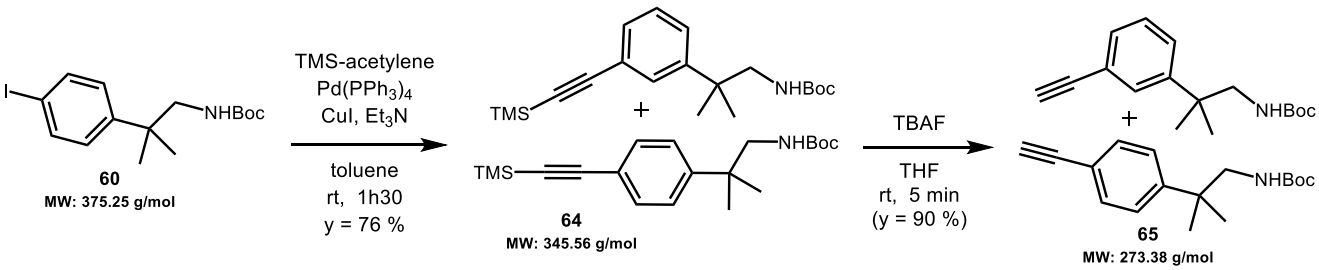
^{13}C NMR shifts extrapolated from the HSQC experiment:

δ = 137.4 (*para* <u>*CH*</u>*-C-I Ar*), 135.6 (*meta CH Ar*), 128.4 (*para CH Ar*), 125.4 (*meta CH Ar*), 51.7 (*CH_2*), 28.3 (*tBu*), 26.3 (*CH_3*).

HSQC:



(65) Synthesis and characterization of ***tert*-butyl (2-(4-ethynylphenyl)-2-methylpropyl)carbamate (65)**:



***tert*-butyl (2-(4-iodophenyl)-2-methylpropyl)carbamate 60** (123 mg, 0.33 mmol, 1 eq), $Pd(Ph_3)_4$ (16 mg, 0.02 mmol, 0.05 eq) and CuI (13 mg, 0.07 mmol, 0.2 eq) were dissolved in toluene (850 µL), under Ar atmosphere. TMS-acetylene (60 µL, 0.42 mmol, 1.3 eq) and Et_3N (60 µL, 0.43 mmol, 1.3 eq) were added and the reaction mixture was stirred at room temperature for 1h30 until TLC showed completion, before being concentrated. The crude was purified by automatic chromatography (Biotage Sfär 25: nHex/EtOAc gradient 0 % to 20

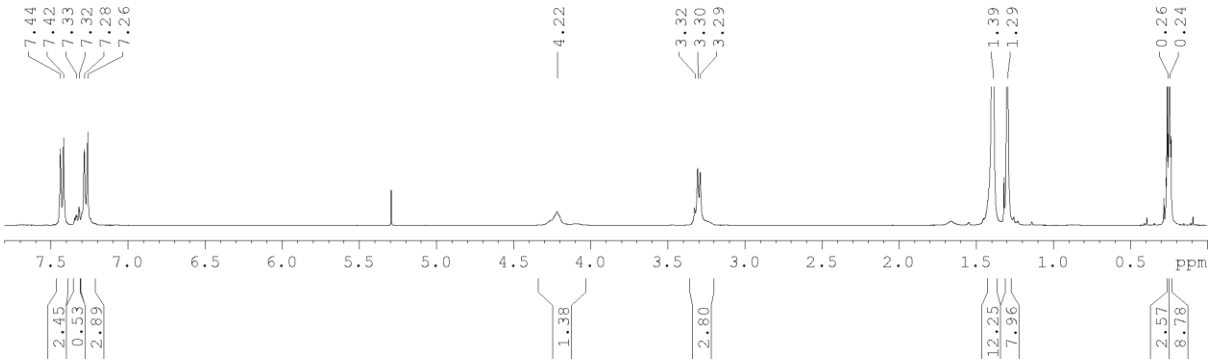
%) affording intermediate **64** (87 mg, 0.25 mmol, y = 76 %) as a regioisomeric mixture (*para*/*meta* 71:29), as seen by ^{1}H NMR (*para* signals at δ = 7.43, 7.27 ppm, *meta* at 7.33 ppm). TLC R_f (nHex/EtOAc: 8/2): 0.67. MS (ESI) calculated for $C_{20}H_{31}NO_2Si$ $[M + Na]^+$ *m/z*: 368.20; found: 368.20.

To ***tert*-butyl (2-methyl-2-(4-((trimethylsilyl)ethynyl)phenyl)propyl)carbamate 64** (60 mg, 0.17 mmol, 1 eq) in THF (500 µL) was added a 1 M solution of TBAF in THF (200 µL, 0.20 mmol, 1.2 eq). The brown reaction mixture was stirred at room temperature for 5 min until TLC showed completion, before being concentrated. The crude was dissolved in DCM and washed with a 1 M aqueous HCl solution. The organic phase was dried over Na_2SO_4 and concentrated to crude **65** (43 mg, max yield: 90 %), used for the next step without purification. TLC R_f (nHex/EtOAc: 8/2): 0.59. MS (ESI) calculated for $C_{17}H_{23}NO_2$ $[M + Na]^+$ *m/z*: 296.16; found: 296.10.

Crude *tert*-butyl (2-methyl-2-(4-((trimethylsilyl)ethynyl)phenyl)propyl)carbamate **64**: *para*/*meta* ratio (71:29)

^{1}H NMR (400 MHz, $CDCl_3$):

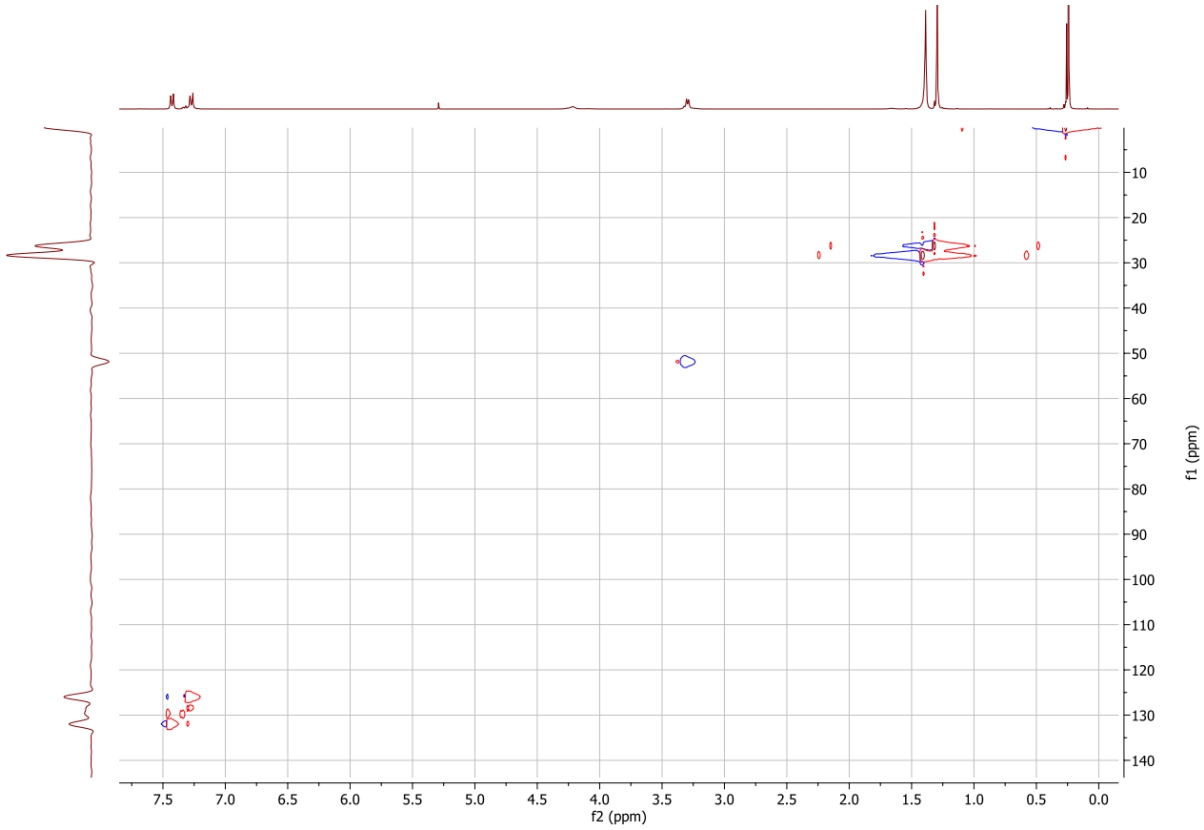
δ = 7.43 (d, J = 8.4 Hz, 2H, *para ≡C-C-CH Ar*), 7.33 (m, 1H, *meta CH Ar*), 7.27 (mult., 4H, *para CH Ar + meta CH Ar*), 4.22 (bs, 1H, *NH*), 3.30 (d, 2H, *para* CH_2), 3.31 (d, 2H, *meta* CH_2), 1.39 (bs, 9H, *tBu*), 1.28 (s, 6H, CH_3), 0.26 (s, 9H, *meta Si-*CH_3), 0.24(s, 9H, *para Si-*CH_3).



^{13}C NMR shifts extrapolated from the HSQC experiment:

δ = 132.0 (*para CH-C-C≡ Ar*), 129.8, 129.6, 128.4 (*meta CH Ar*), 126.0 (*para CH Ar*), 51.9 (CH_2), 28.3 (*tBu*), 26.3 (CH_3), 0.2 (Si-CH_3).

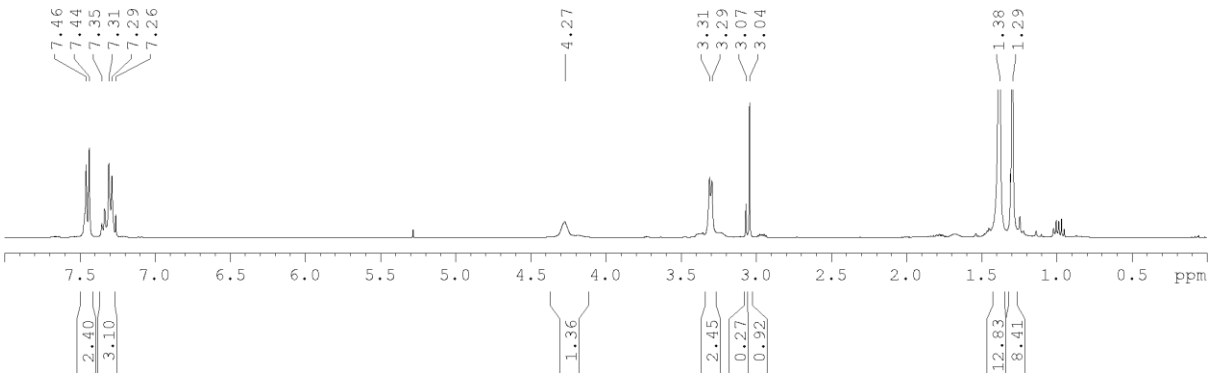
HSQC:



Crude *tert*-butyl (2-(4-ethynylphenyl)-2-methylpropyl)carbamate **65**: *para*/*meta* ratio (71:29)

^{1}H NMR (400 MHz, $CDCl_3$):

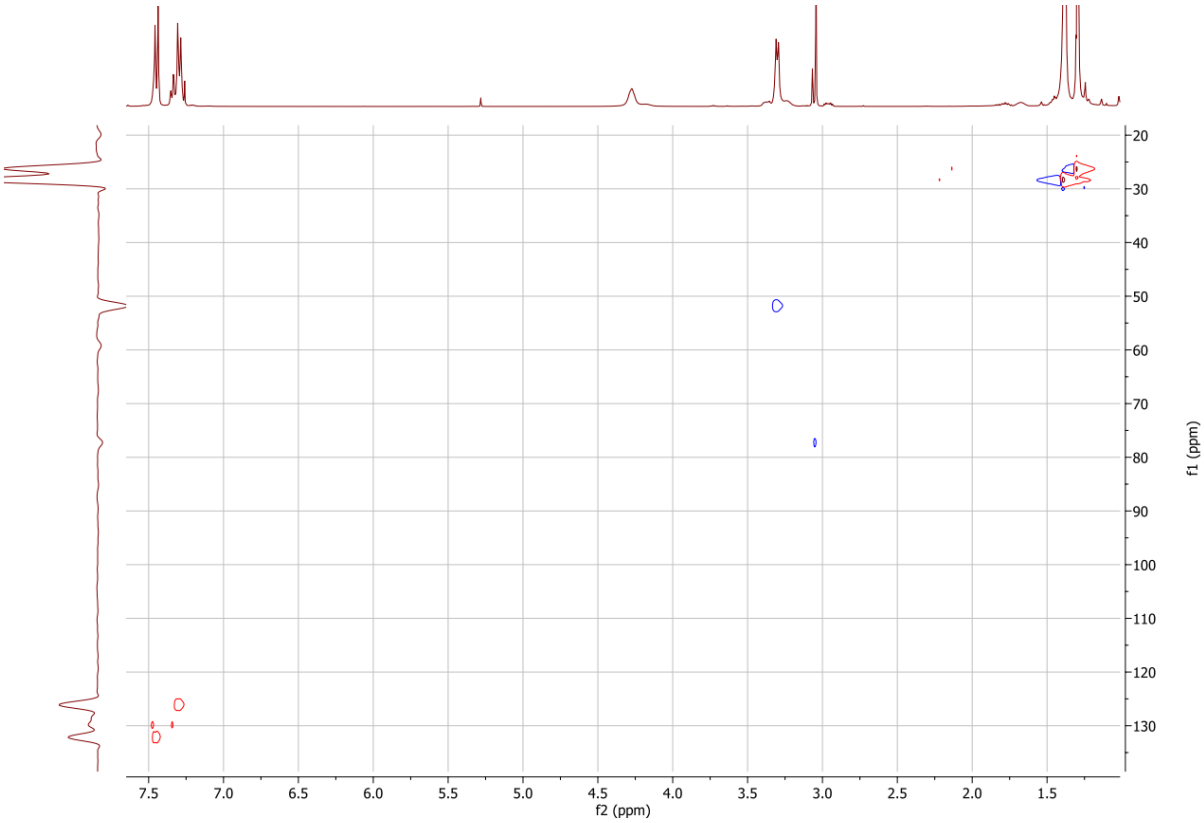
δ = 7.45 (d, J = 8.4 Hz, 2H, *para* <u>*CH*</u>*-C-C≡ Ar*), 7.35 - 7.26 (mult., 5H, 2 x *para CH Ar* + 3 x *meta CH Ar*), 4.27 (bs, 1H, *NH*), 3.30 (d, J = 6.2 Hz, 2H, CH_2), 3.07 (s, 1H, *meta ≡CH*), 3.04 (s, 1H, *para ≡CH*), 1.38 (bs, 9H, *tBu*), 1.29 (s, 6H, CH_3).



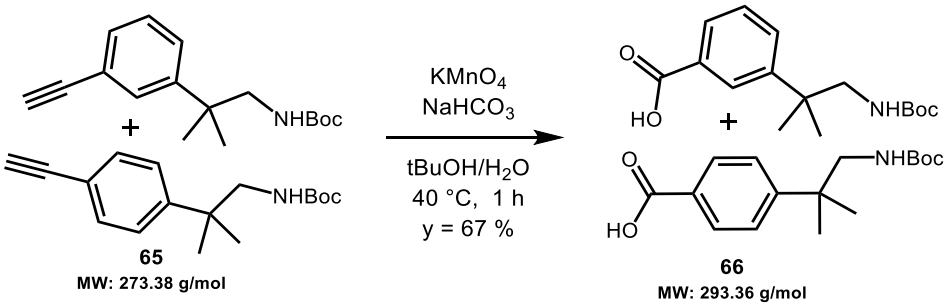
^{13}C NMR shifts extrapolated from the HSQC experiment:

δ = 132.1 (*para* <u>*CH*</u>*-C-C≡ Ar*), 129.9, 129.9, 128.5 (*meta CH Ar*), 126.1 (*para CH Ar*), 77.3 (*≡CH*), 51.7 (CH_2), 28.3 (*tBu*), 26.2 (CH_3).

HSQC:



(66) Synthesis and characterization of **4-(1-((tert-butoxycarbonyl)amino)-2-methylpropan-2-yl) benzoic acid (66)**:

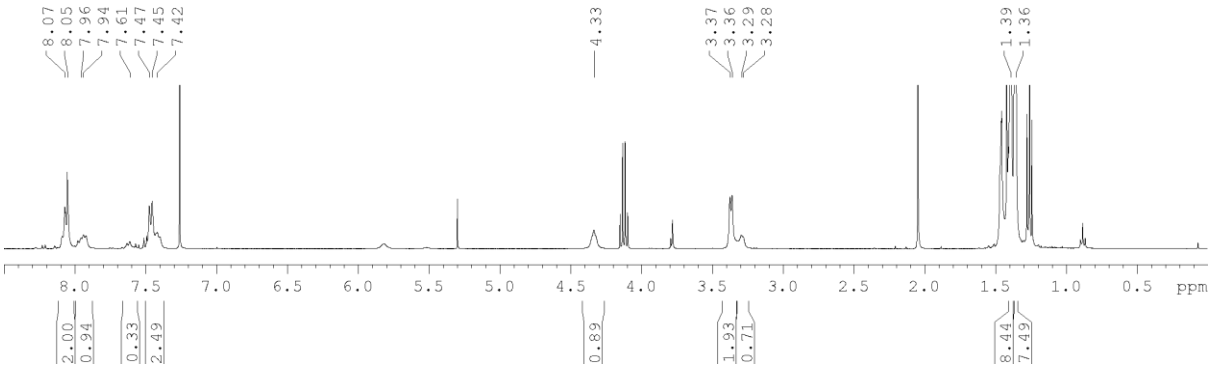


***tert*-butyl (2-(4-ethynylphenyl)-2-methylpropyl)carbamate 65** (246 mg, 0.90 mmol, 1 eq) was dissolved in tBuOH (30 mL). Separately, $KMnO_4$ (426 mg, 2.70 mmol, 3 eq) was dissolved in water (30 mL). Separately, $NaHCO_3$ (250 mg, 3.0 mmol, 3.3 eq) was dissolved in water (30 mL). The $NaHCO_3$ solution, followed by the $KMnO_4$ solution were added to the dissolved alkyne and stirred at 40 °C for 1 h until TLC showed completion. The reaction mixture was quenched by stirring with MeOH, then was filtered and concentrated. The crude was redissolved in DCM and acidified water (to pH 2, using 1M HCl). The organic phase was washed with brine, dried over Na_2SO_4, and concentrated. The crude was purified by automatic

chromatography (Biotage Sfär 10: nHex/EtOAc gradient 0 % to 60 %; 0.01% formic acid) affording **66** (177 mg, 0.60 mmol, y = 67 %). TLC R_f (nHex/EtOAc: 8/2; 0.01% formic acid): 0.16. MS (ESI) calculated for $C_{16}H_{23}NO_4$ $[M + Na]^+$ *m/z*: 316.15; found: 316.31.

^{1}H NMR (400 MHz, $CDCl_3$) *para/meta* ratio (66:34):

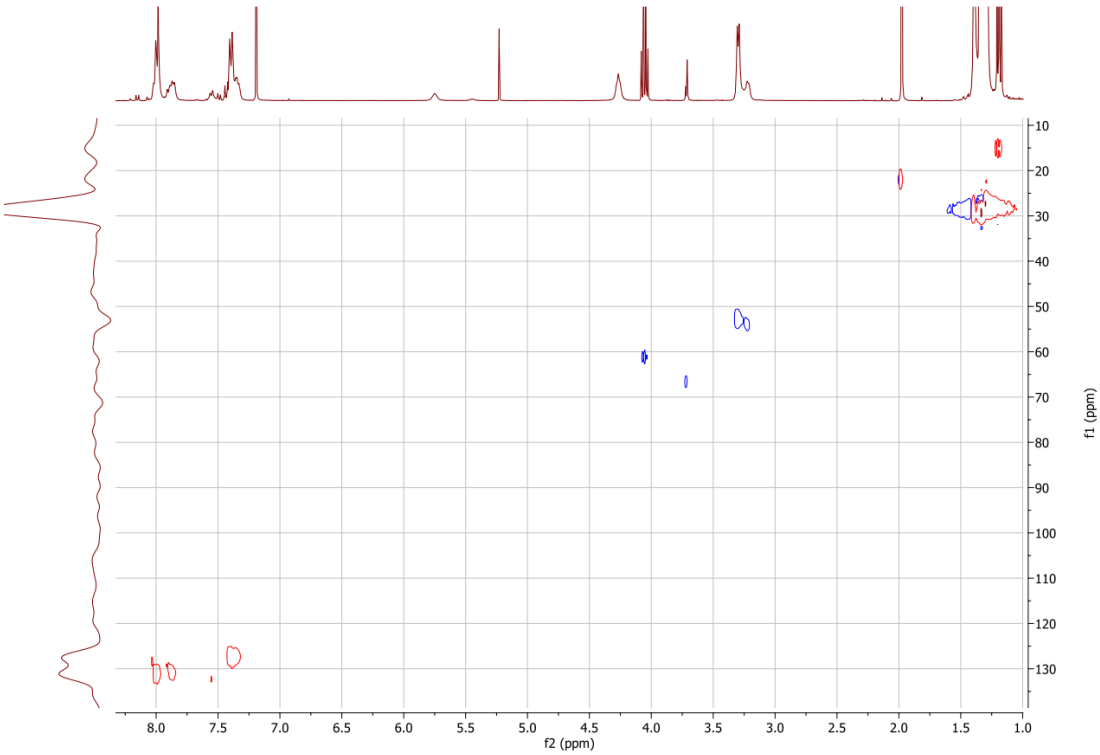
δ = 8.06 (d, 2H, *para* <u>*CH*</u>*-C-C=O Ar*), 7.95 (m, 1H, *meta CH Ar*), 7.61 (m, 1H, *meta CH Ar*), 7.46 (m, 2H, *para CH Ar*), 7.42 (m, 1H, meta *CH Ar*), 4.33 (bs, 1H, *NH*), 3.37 (d, J = 6.1 Hz, 2H, CH_2), 3.29 (d, 2H, CH_2), 1.39 (bs, 9H, *tBu*), 1.36 (s, 6H, CH_3).



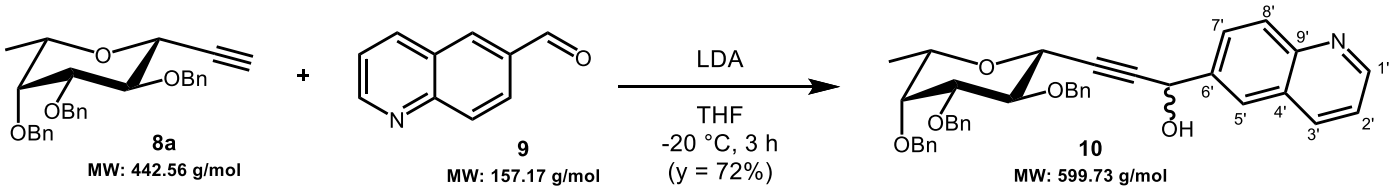
^{13}C NMR shifts extrapolated from the HSQC experiment:

δ = 132.3 (*meta CH Ar*), 131.1 (*para* <u>*CH*</u>*-C-I Ar*), 130.9 (*meta CH Ar*), 127.3 (*para CH Ar*), 54.0 (*meta* CH_2), 52.7 (*para* CH_2), 29.3 (*tBu*), 27.4 (CH_3).

HSQC:



(10) Synthesis and characterization of **1-(quinolin-6-yl)-3-(2,3,4-tri-O-benzyl β-L-fucopyranosyl)prop-2-yn-1-ol (10)** following the procedure of Dondoni and co-workers:[192]

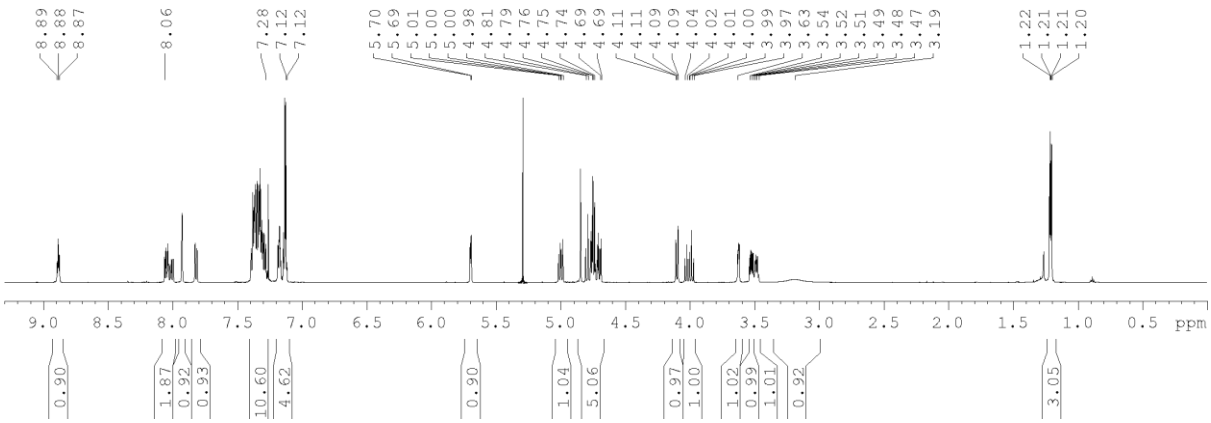


A solution of **(2,3,4-tri-O-benzyl β-L-fucopyranosyl) acetylene 8a** (83 mg, 0.19 mmol, 1.2 eq) in dry THF (2.5 mL) was cooled to -20 °C under Ar atmosphere. A freshly made 0.3 M solution of LDA in THF (850 µL, 0.26 mmol, 1.6 eq) was added to the solution and left to stir for 10 min while **quinoline-6-carbaldehyde 9** (26 mg, 0.16 mmol, 1 eq) was dissolved in dry THF (1.5 mL), then added to the reaction. The mixture was stirred at -20 °C for 2 h, then at -10 °C for 1h until TLC showed completion. The reaction was quenched with water and extracted with EtOAc. The organic phase was washed with water and dried over Na_2SO_4. The crude product was purified by automatic chromatography (Biotage Sfär 10: nHex/EtOAc gradient from 15% to 100%) affording the diastereomeric mixture **10** in a 1:1 ratio (71 mg, 0.12 mmol, 72 %), as seen by ^{1}H NMR (clearly doubled signal for *CH-OH* and *H-2* at δ = 5.70, 5.69 and 4.02, 3.99 ppm). TLC R_f (DCM/MeOH: 97/3): 0.30. MS (ESI) calculated for $C_{39}H_{37}NO_5$ $[M + Na]^+$ *m/z*: 622.26; found: 622.60.

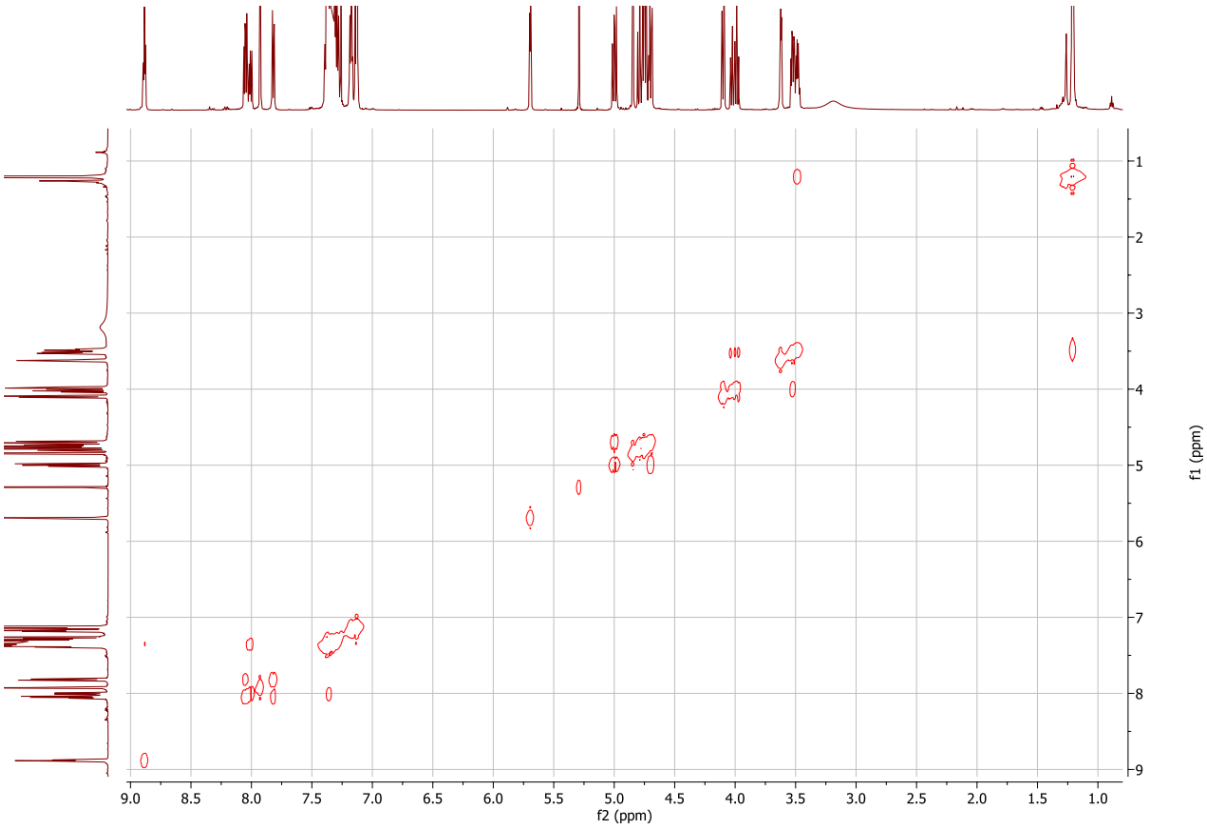
^{1}H NMR (400 MHz, $CDCl_3$): mixture of **10** diastereoisomers (ratio 1:1)

Product **10**:

δ = 8.88 (m, 1H, *H-1'*), 8.06 - 8.00 (mult., $J_{7'-8'}$ = 8.8 Hz, $J_{7'-5'}$ = 5.6 Hz, 2H, *H-3'* + *H-7'*), 7.93 (bs, 1H, *H-5'*), 7.82 (dd, $J_{8'-7'}$ = 8.8 Hz, $J_{8'-5'}$ = 2.0 Hz, 1H, *H-8'*), 7.39 - 7.28 (mult., 11H, *CH Ar* + *H-2'*), 7.19 - 7.12 (mult., 5H, *CH Ar*), 5.70, 5.69 (mult., 1H, *HC-OH*), 5.01 - 4.98 (dd, 1H, *CH_2 Ar*), 4.84 - 4.69 (mult., 5H, *CH_2 Ar*), 4.10 (d, J_{1-2} = 9.6 Hz 1H, *H-1*), 4.02, 3.99 (t, J_{2-1} = J_{2-3} = 9.5 Hz, 1H, *H-2*), 3.63 (m, J_{4-3} = 2.9 Hz, 1H, *H-4*), 3.52 (m, J_{3-2} = 9.4 Hz, J_{3-4} = 2.8 Hz, 1H, *H-3*), 3.48 (m, J_{5-CH3} = 6.5 Hz, 1H, H-*5*), 3.19 (bs, 1H, *OH*), 1.21 (d, J_{CH3-5} = 6.4 Hz, 3H, *CH_3*).



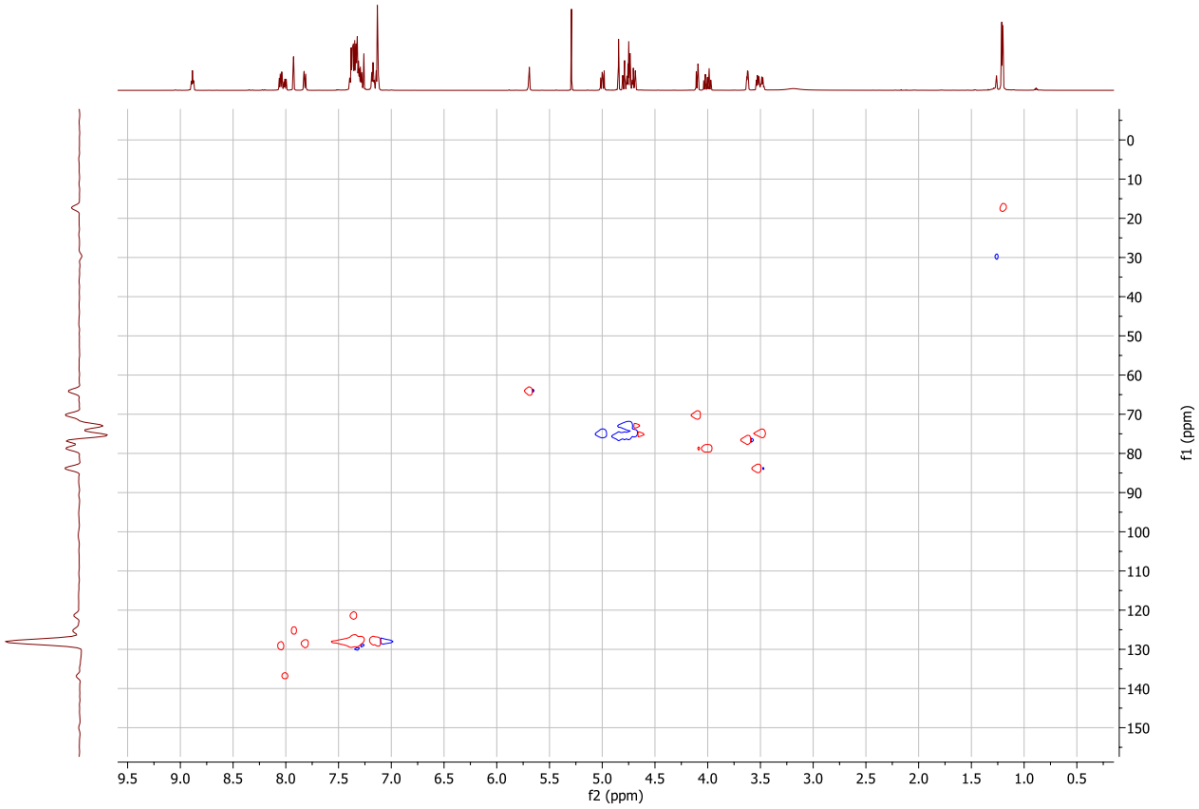
COSY:



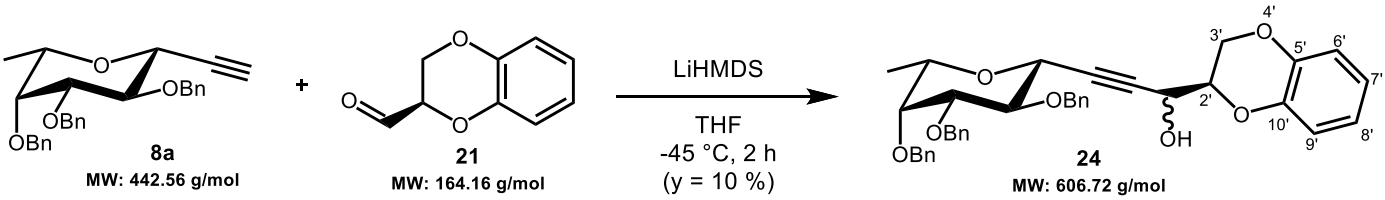
^{13}C NMR shifts extrapolated from the HSQC experiment:

δ = 149.8 (*C1'*), 136.8 (*C3'*), 129.0 (*C7'*), 128.5 (*C8'*), 127.9 (*CH Ar*), 125.1 (*C5'*), 121.4 (*C2'*), 83.8 (*C3*), 78.7 (*C2*), 76.5 (*C4*), 74.8 (*C5*), 75.5, 74.9, 72.9 (CH_2 *OBn*), 70.2 (*C1*), 64.0 (*CH-OH*), 17.2 (CH_3).

HSQC:



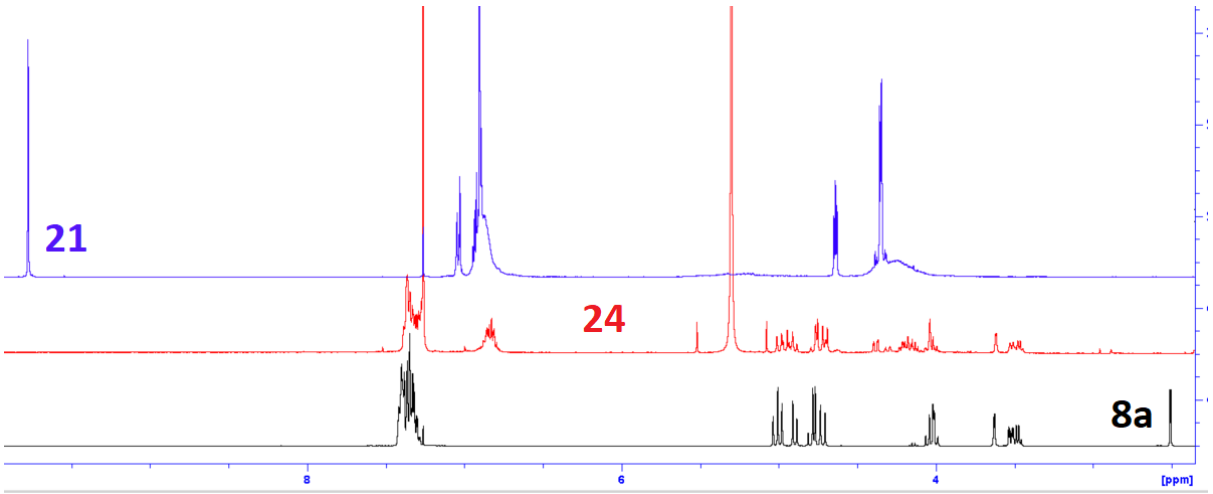
(24) Synthesis and characterization of **1-((R)-1,4-benzodioxan-2-yl)-3-(2,3,4-tri-O-benzyl β-L-fucopyranosyl)prop-2-yn-1-ol (24)** following the procedure of Dondoni and co-workers:[192]



A solution of **(2,3,4-tri-O-benzyl β-L-fucopyranosyl) acetylene 8a** (20 mg, 0.045 mmol, 1.0 eq) in freshly distilled THF (250 µL) was cooled to -20 °C under Ar atmosphere. A 1 M solution of LiHMDS in THF (113 µL, 0.113 mmol, 2.5 eq) was added to the solution and left to stir for 1 h while **(R)-1,4-benzodioxane-2-carboxaldehyde 21** (12 mg, 0.073 mmol, 1.6 eq) was dissolved in dry THF (250 µL) under Ar atmosphere. The alkyne solution was cooled -45 °C prior to the addition of the aldehyde solution. The mixture was stirred at -10 °C for 2 h until TLC showed full conversion of the aldehyde to a new product (a portion of **8a** remained unreacted), before being quenched with water and extracted with EtOAc. The organic phase was washed with water and dried over Na_2SO_4. The crude product was purified by flash chromatography (nHex/EtOAc: 6/4), affording unreacted alkyne **8a** (13 mg, 0.029 mmol, 65 %) and product **24** (1 mg, 0.002 mmol, 10 % of the 0.016 mmol of reacted material), as seen by ^{1}H NMR (see

below). A second fraction (2 mg, 0.003 mmol, 21 %) contained a mixture of product **24** with stronger signals for the minor stereoisomer at δ = 4.30 ppm (weaker at 4.38 ppm). TLC R_f (nHex/EtOAc: 6/4): 0.42.

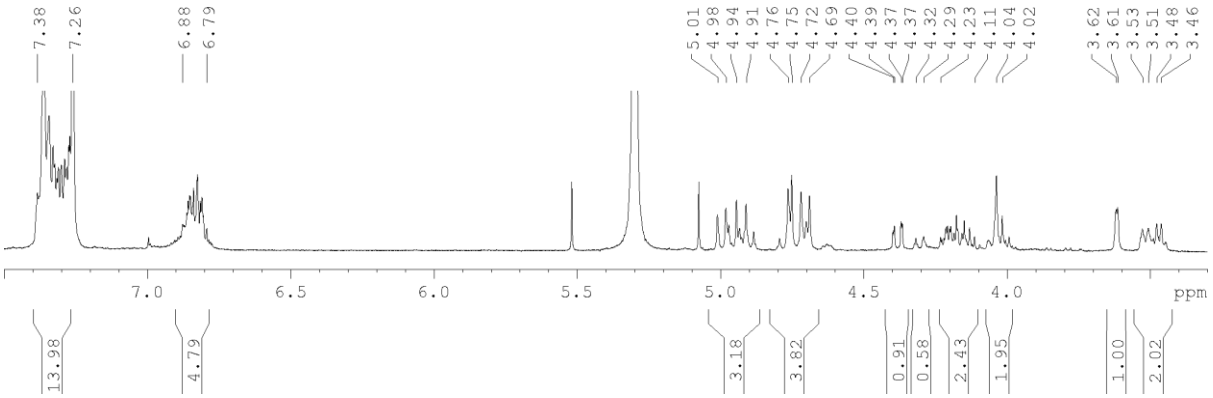
^{1}H NMR (400 MHz, $CDCl_3$): comparison of starting materials and product **24**



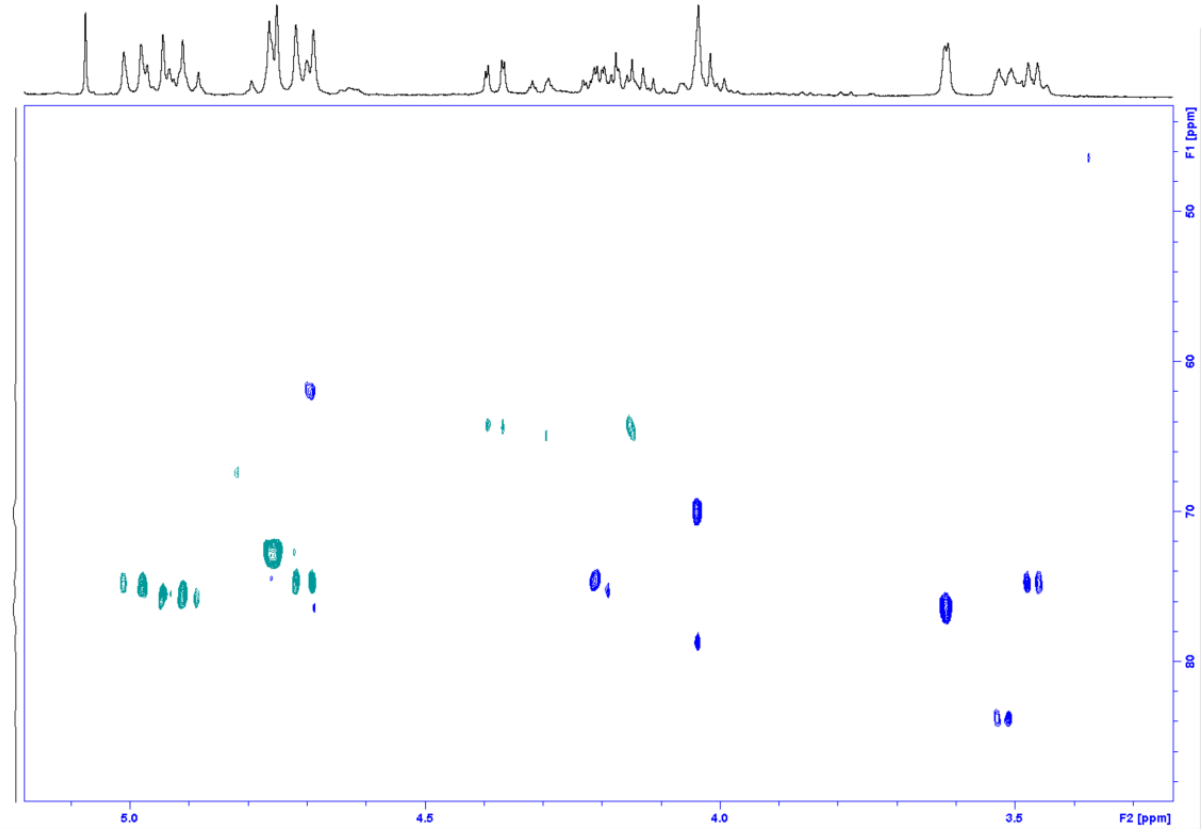
δ = 9.77 (**21**, s, 1H, aldehyde), 2.51 (**8a**, s, 1H, free alkyne).

^{1}H NMR (400 MHz, $CDCl_3$): Product **24** (zoom to relevant region, 7.5 - 3.3 ppm)

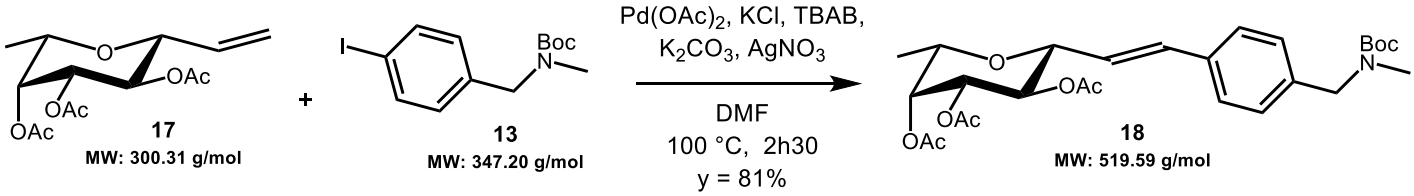
δ = 7.38 - 7.26 (mult., 15H, *OBn Ar*), 6.88 - 6.79 (mult., 4H, *CH Ar*), 5.01 - 4.91 (mult., 3H, *CH_2 OBn*), 4.76 - 4.69 (mult., 4H, *CH_2 OBn* + *H2'*), 4.38 (dd, J = 11.0 Hz, J' = 1.8 Hz, 1H, *H-3'*), **4.30 (d, J = 11.0 Hz,** second stereoisomer, ratio 2:3, 1H, *H-3'***)**, 4.23 - 4.11 (mult., 2H, *H-3''* + *HC-OH*), 4.04 (mult., 2H, *H-1* + *H-2*), 3.61 (d, 1H, *H-4*), 3.49 (mult., 2H, *H-3* + *H-5*), 1.19 (d, 3H, *CH_3*).



HSQC:



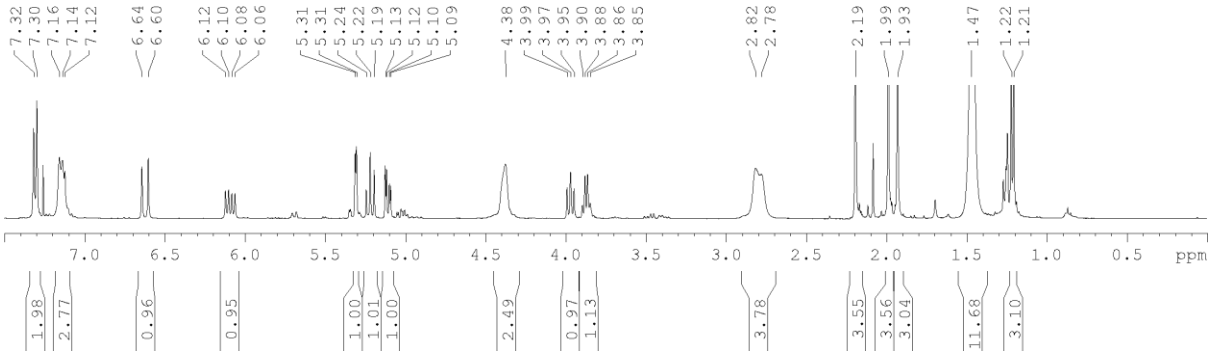
(18) Synthesis and characterization of ***tert*-butyl (*E*)-(4-(2,3,4-tri-O-acetyl β-L-fucopyranosylvinyl)benzyl) (methyl) carbamate (18)** following the procedure of Kondor and co-workers:[191]



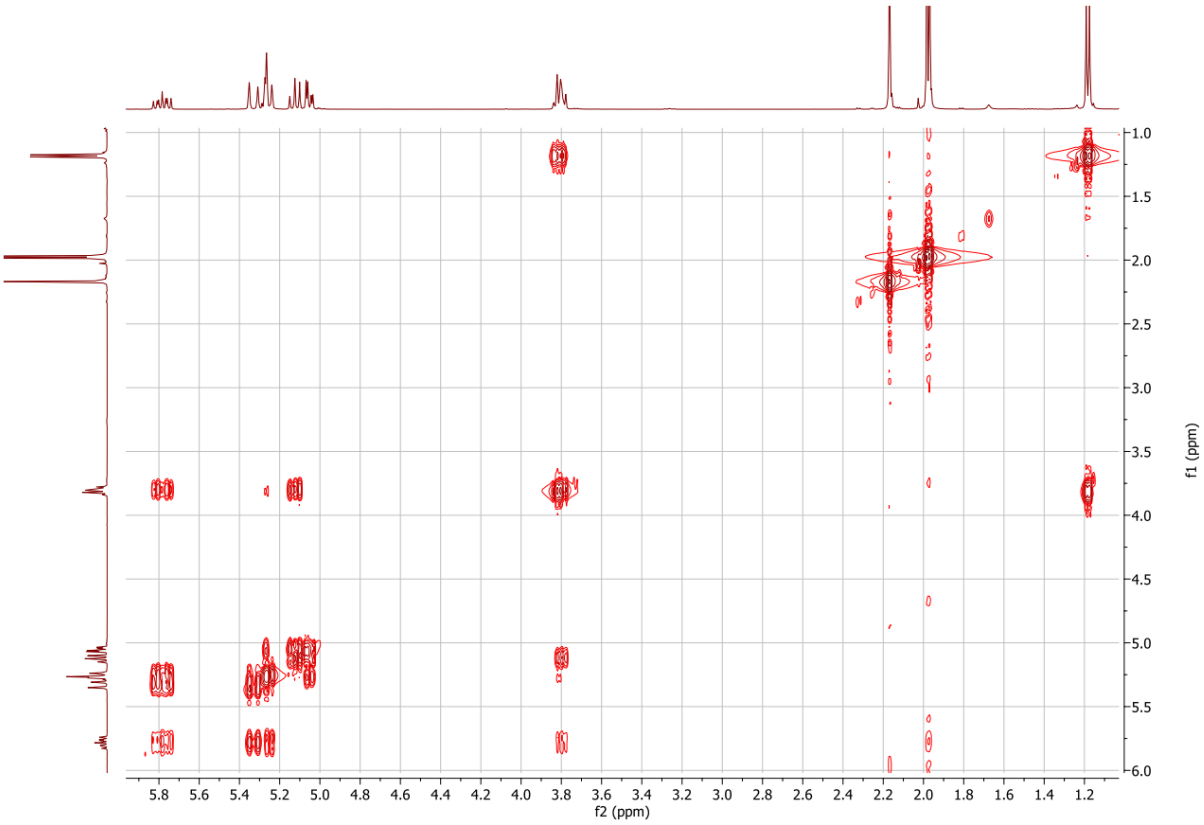
The pure *para*-iodinated isomer of **13** (synthesized by Method 2) ***tert*-butyl (4-iodobenzyl)(methyl)carbamate** (32 mg, 0.092 mmol 1.1 eq) and **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) ethene 17** (25 mg, 0.083 mmol, 1 eq) were dissolved in dry DMF (1.7 mL) under Ar atmosphere. $Pd(OAc)_2$ (2 mg, 0.009 mmol, 0.11 eq), KCl (7 mg, 0.094 mmol, 1.1 eq), TBAB (52 mg, 0.161 mmol, 1.9 eq), K_2CO_3 (20 mg, 0.145 mmol, 1.7 eq) and $AgNO_3$ (3 mg, 0.018 mmol, 0.21 eq) were added to the solution. The reaction mixture was heated to 100 °C for 2h30 before returning to room temperature. The crude was concentrated and purified by automatic chromatography (Biotage Sfär 10: nHex/EtOAc gradient 5 % to 70 %) affording product **18** (35 mg, 0.067 mmol, y = 81 %). TLC R_f (nHex/EtOAc: 7/3): 0.38. MS (ESI) calculated for $C_{27}H_{37}NO_9$ $[M + Na]^+$ m/z: 542.24; found: 542.39.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.31 (d, J = 8.0 Hz, 2H, *CH=CH-C-CH Ar*), 7.14 (d, 2H, *CH Ar*), 6.62 (d, J_{trans} = 15.8 Hz, 1H, *C1-CH=CH*), 6.09 (dd, J_{trans} = 15.8 Hz, J_{CH-1} = 7.8 Hz, 1H, *C1-CH=CH*), 5.31 (d, J_{4-3} = 3.5 Hz, 1H, *H-4*), 5.22 (dd, J_{2-1} = 9.8 Hz, J_{2-3} = 10.0 Hz, 1H, *H-2*), 5.11 (dd, J_{3-2} = 10.2 Hz, J_{3-4} = 3.4 Hz, 1H, *H-3*), 4.38 (s, 2H, CH_2), 3.97 (dd, J_{1-2} = J_{CH-1} = 8.7 Hz, 1H, *H-1*), 3.87 (qd, J_{5-4} = 0.8 Hz, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 2.80 (bd, 3H, $N-CH_3$), 2.19 (s, 3H, *OAc*), 1.99 (s, 3H, *OAc*), 1.93 (s, 3H, *OAc*), 1.47 (bs, 9H, *tBu*), 1.21 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).

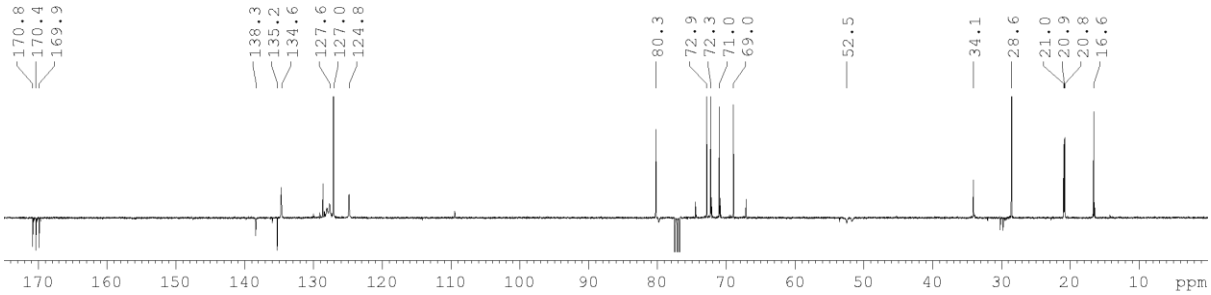


COSY:

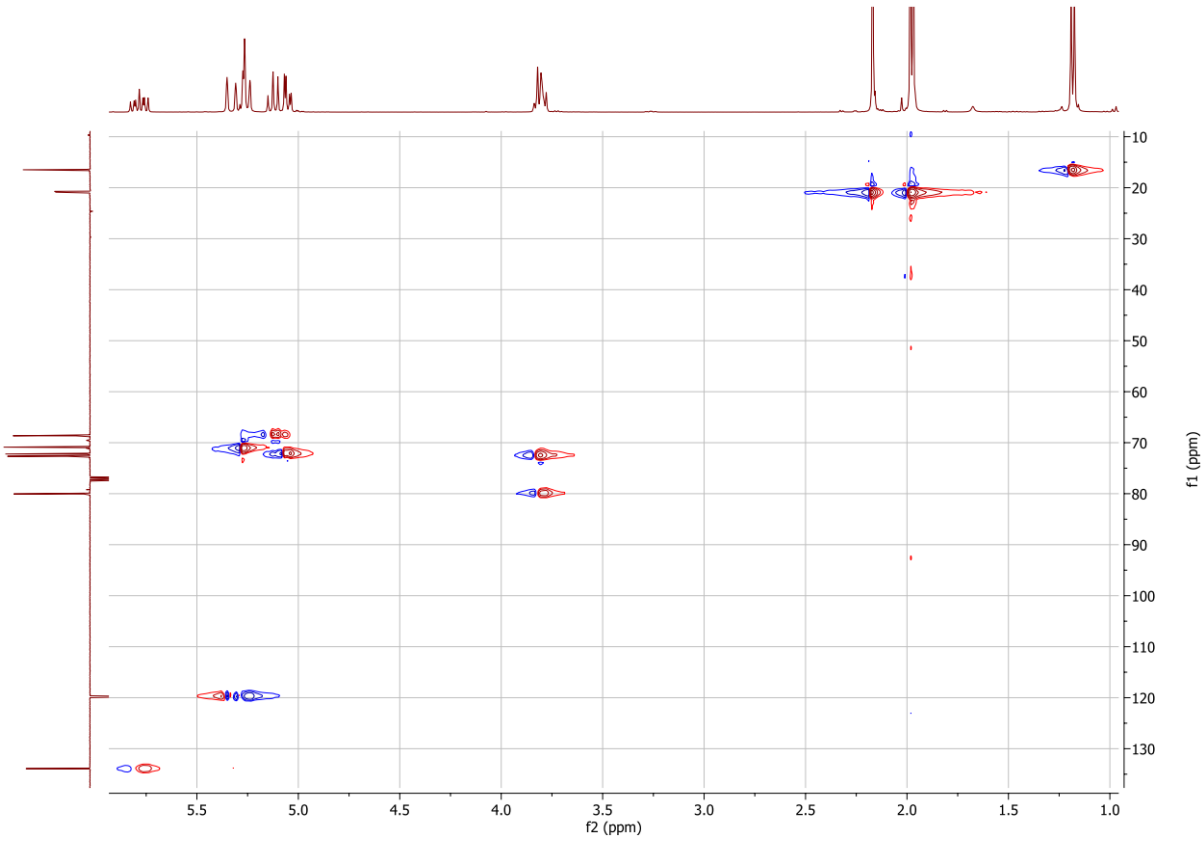


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

δ = 170.8, 170.4, 169.6 (*C=O*), 138.3 (*CH=CH-C-CH Ar*), 135.2 (*C Ar*), 134.6 (*C1-CH=CH*), 127.6 (*CH=CH-C-CH Ar*), 127.0 (*CH Ar*), 124.8 (*C1-CH=CH*), 80.3 (*C1*), 72.9 (*C5*), 72.3 (*C3*), 71.0 (*C4*), 69.0 (*C2*), 52.5 (CH_2), 34.1 (CH_3), 28.6 (CH_3 *tBu*), 21.0, 20.9, 20.8 (CH_3 *OAc*), 16.6 (*C6*).



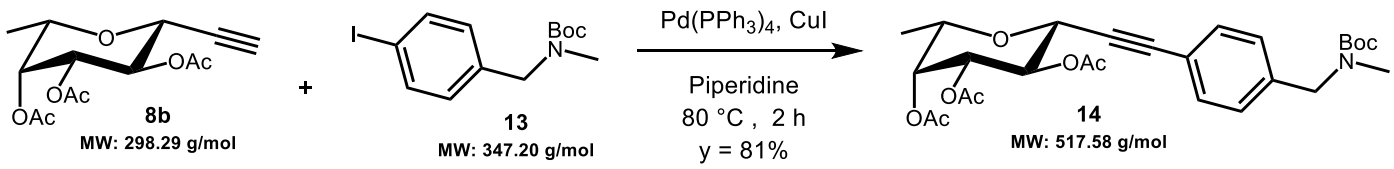
HSQC:



General procedure for Sonogashira coupling for β-C-fucosylacetylenes, from Lowary and co-workers:[181]

The **iodinated fragment** (1.1 eq) was dissolved in piperidine (concentration: 0.2 M) and added to a flask containing **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) acetylene 14** (1 eq), $Pd(PPh_3)_4$ (0.05 eq) and CuI (0.10 eq) under Ar atmosphere. The reaction mixture was heated to 80 °C for 2 h until TLC showed completion, before returning to room temperature. The crude was concentrated and re-dissolved in pyridine (1 - 2 mL) and Ac_2O (1 - 2 mL) and stirred overnight at room temperature to re-acetylate the eventual deprotected positions. The reaction mixture was cooled to 0 °C and treated with ice-cold MeOH while stirring, then diluted with toluene for co-evaporation of pyridine. Pyridine was also removed by diluting the crude with EtOAc and repeatedly washing with a 0.02 M HCl solution. The organic phase was dried over Na_2SO_4 and concentrated. The crude was purified by flash or automatic chromatography (nHex/EtOAc: 6/4) affording the coupled product.

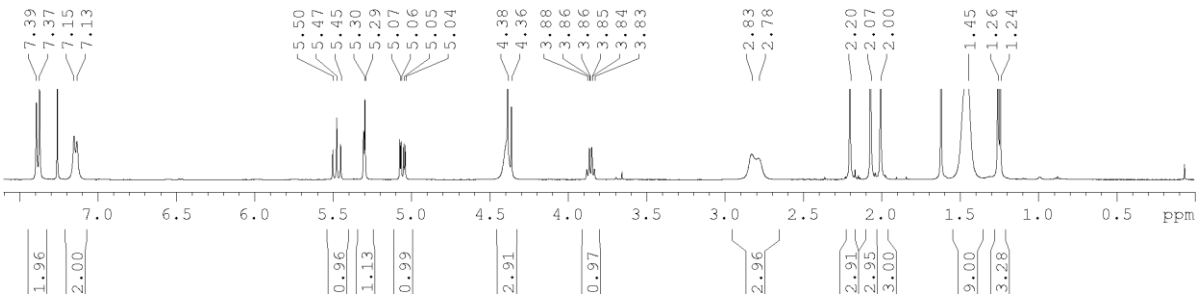
(14) Synthesis and characterization of ***tert*-butyl (4-(2,3,4-tri-O-acetyl β-L-fucopyranosylethynyl)benzyl) (methyl) carbamate (14)**:



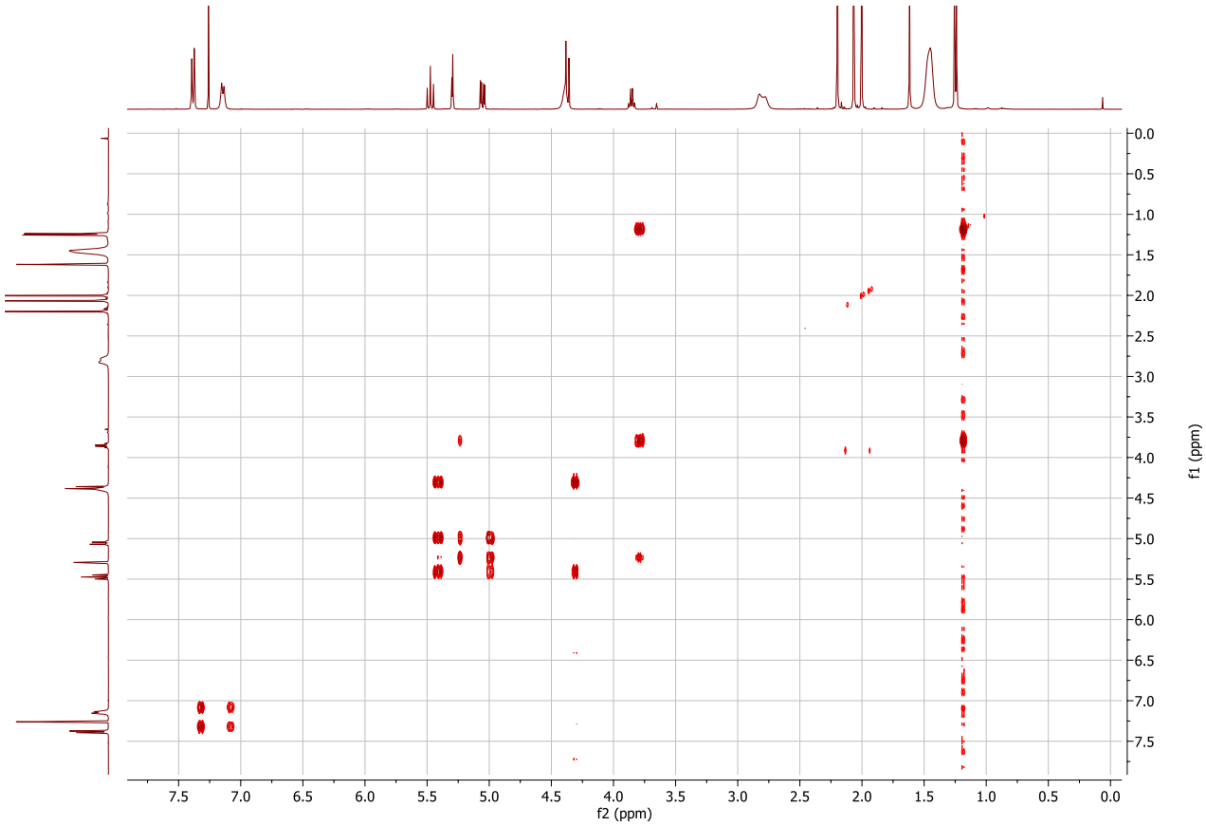
The pure *para*-iodinated isomer of **13** (synthesized by Method 2) ***tert*-butyl (4-iodobenzyl)(methyl)carbamate** (0.37 mmol) and **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) acetylene 8b** (0.34 mmol) were coupled following the aforementioned procedure to afford **14** (0.27 mmol, y = 81 %). TLC R_f (nHex/EtOAc: 6/4): 0.55. MS (ESI) calculated for $C_{27}H_{35}NO_9$ $[M + Na]^+$ *m/z*: 540.22; found: 540.24. $[\alpha]_D^{17}$ = -14.0 ($CHCl_3$, c 1).

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.38 (d, J = 8.3 Hz, 2H, *≡C-C-CH Ar*), 7.14 (d, J = 7.8 Hz, 2H, *CH Ar*), 5.47 (dd, $J_{2-1} = J_{2-3}$ = 10.0 Hz, 1H, *H-2*), 5.30 (dd, J_{4-3} = 3.4 Hz, J_{4-5} = 1.2 Hz, 1H, *H-4*), 5.05 (dd, J_{3-2} = 10.2 Hz, J_{3-4} = 3.4 Hz, 1H, *H-3*), 4.39 (s, 2H, CH_2), 4.37 (d, J_{1-2} = 9.8 Hz, 1H, *H-1*), 3.85 (qd, J_{5-4} = 1.2 Hz, J_{5-CH3} = 6.4 Hz, 1H, *H-5*), 2.80 (bd, 3H, *N-CH_3*), 2.20 (s, 3H, *OAc*), 2.07 (s, 3H, *OAc*), 2.00 (s, 3H, *OAc*), 1.45 (bs, 9H, *tBu*), 1.25 (d, J_{CH3-5} = 6.4 Hz, 3H, CH_3).

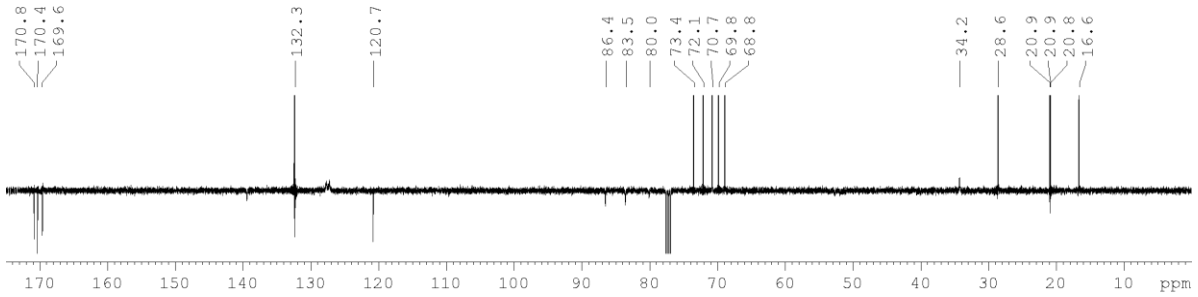


COSY:

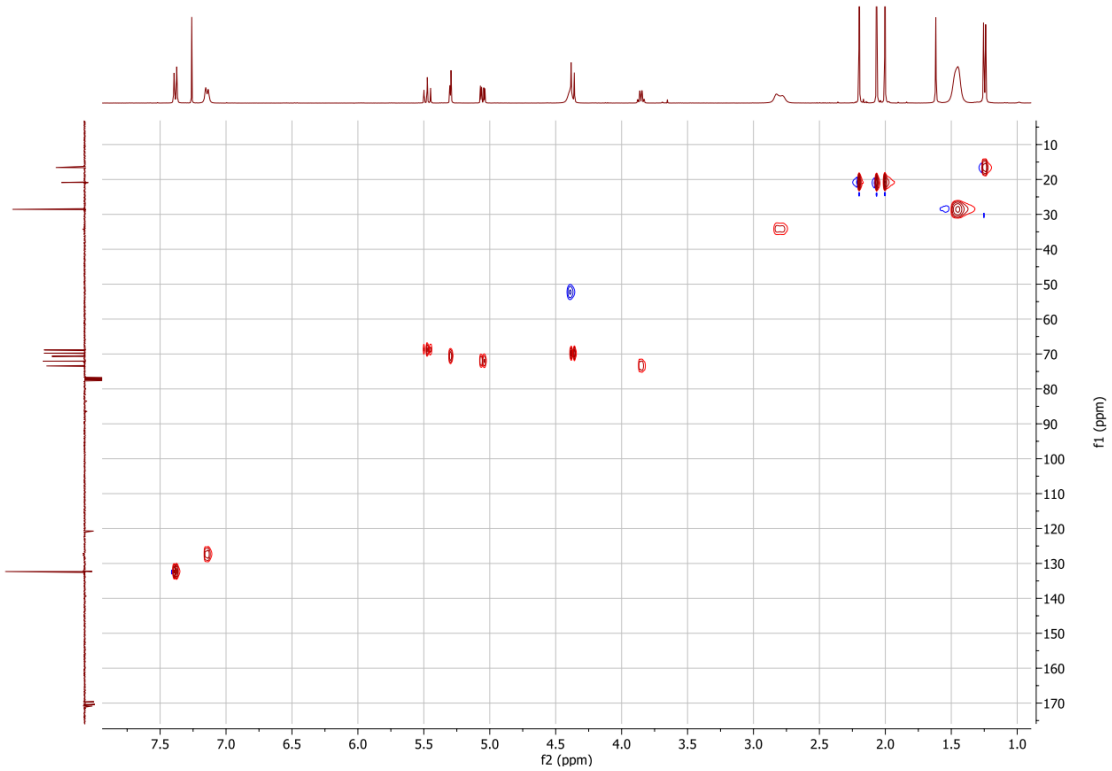


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

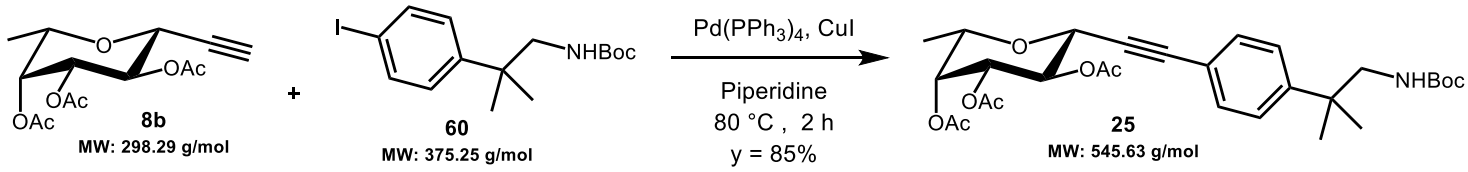
δ = 170.8, 170.4, 169.6 (*C=O*), 132.3 (*≡C-C-<u>CH</u> Ar*), 127.3 (*CH Ar*), 120.7 (*≡C-<u>C</u> Ar*), 86.4 (C1-*<u>C</u>≡*), 83.5 (*≡<u>C</u>-Ar*), 80.0 (*C tBu*), 73.4 (*C5*), 72.1 (*C3*), 70.7 (*C4*), 69.8 (*C1*), 68.8 (*C2*), 52.2 (CH_2), 34.2 (CH_3), 28.6 (CH_3 *tBu*), 20.9, 20.9, 20.8 (CH_3 *OAc*), 16.6 (*C6*).



HSQC:



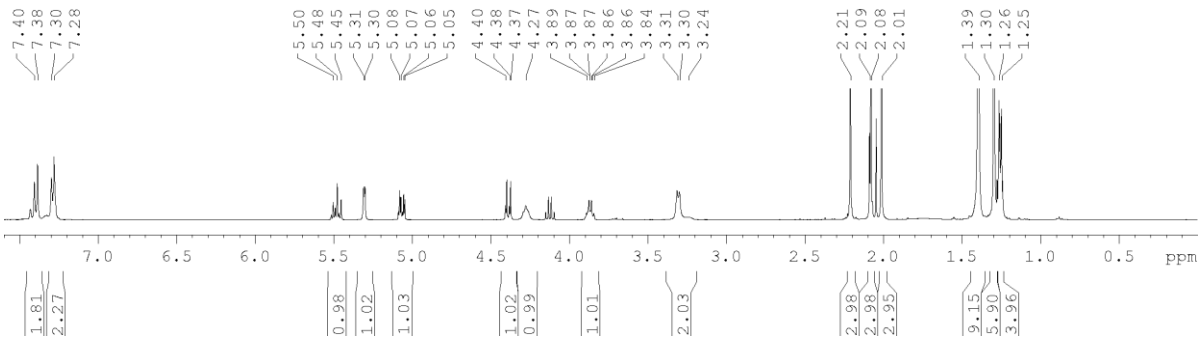
(25) Synthesis and characterization of ***tert*-butyl (2-(4-(2,3,4-tri-O-acetyl β-L-fucopyranosylethynyl)phenyl)-2-methylpropyl)carbamate (25)**:



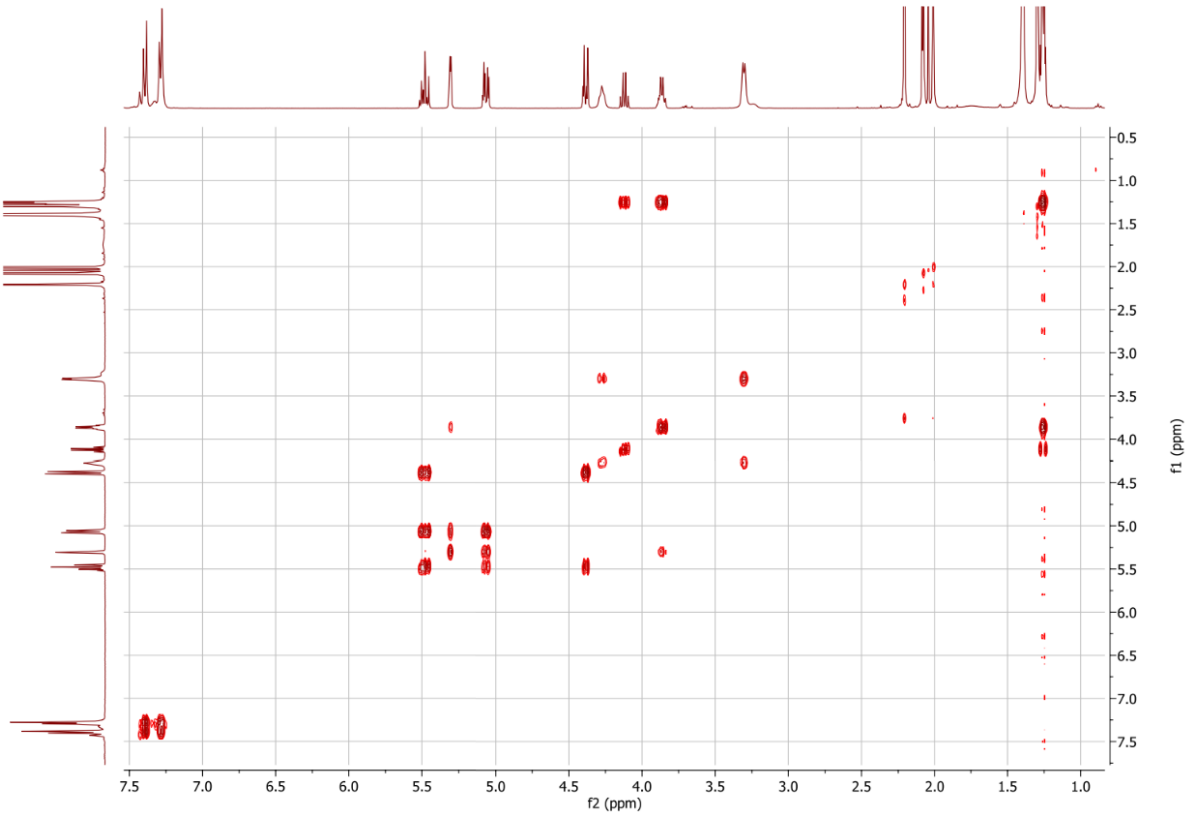
***tert*-butyl (2-(4-iodophenyl)-2-methylpropyl)carbamate 60** (0.139 mmol) and **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) acetylene 8b** (0.128 mmol) were coupled following the aforementioned procedure to afford **25** (0.108 mmol, y = 85 %). TLC R_f (nHex/EtOAc: 6/4): 0.37. MS (ESI) calculated for $C_{29}H_{39}NO_9$ $[M + Na]^+$ *m/z*: 568.25; found: 568.24.

^{1}H NMR (400 MHz, $CDCl_3$): isomeric mixture *para*/*meta* (ratio 87:13)

δ = 7.39 (d, J = 8.5 Hz, 2H, *≡C-C-CH Ar*), 7.29 (m, 2H, *CH Ar*), 5.48 (dd, $J_{2-1} = J_{2-3}$ = 10.0 Hz, 1H, *H-2*), 5.31 (dd, J_{4-3} = 3.4 Hz, J_{4-5} = 1.2 Hz, 1H, *H-4*), 5.06 (dd, J_{3-2} = 10.2 Hz, J_{3-4} = 3.4 Hz, 1H, *H-3*), 4.38 (d, J_{1-2} = 9.8 Hz, 1H, *H-1*), 4.27 (bt, J = 6.4 Hz, 1H, *NH*), 3.86 (qd, J_{5-4} = 1.1 Hz, J_{5-6} = 6.4 Hz, 1H, *H-5*), 3.30 (d, J = 6.4 Hz, 2H, CH_2), 2.21 (s, 3H, *OAc*), 2.08 (s, 3H, *OAc*), 2.01 (s, 3H, *OAc*), 1.39 (bs, 9H, *tBu*), 1.30 (s, 6H, 2xCH_3), 1.26 (d, J_{6-5} = 6.4 Hz, 3H, *C6*).

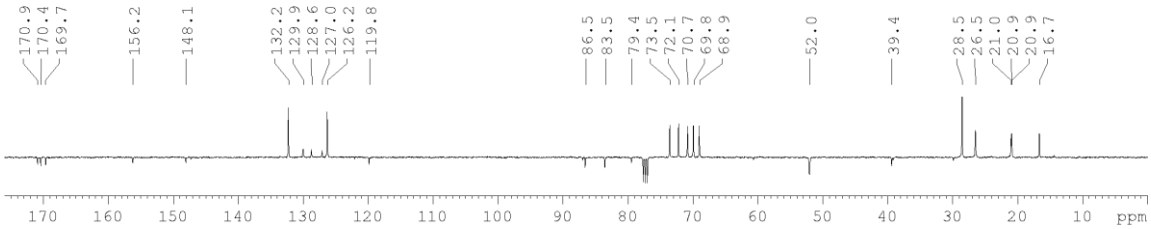


COSY:

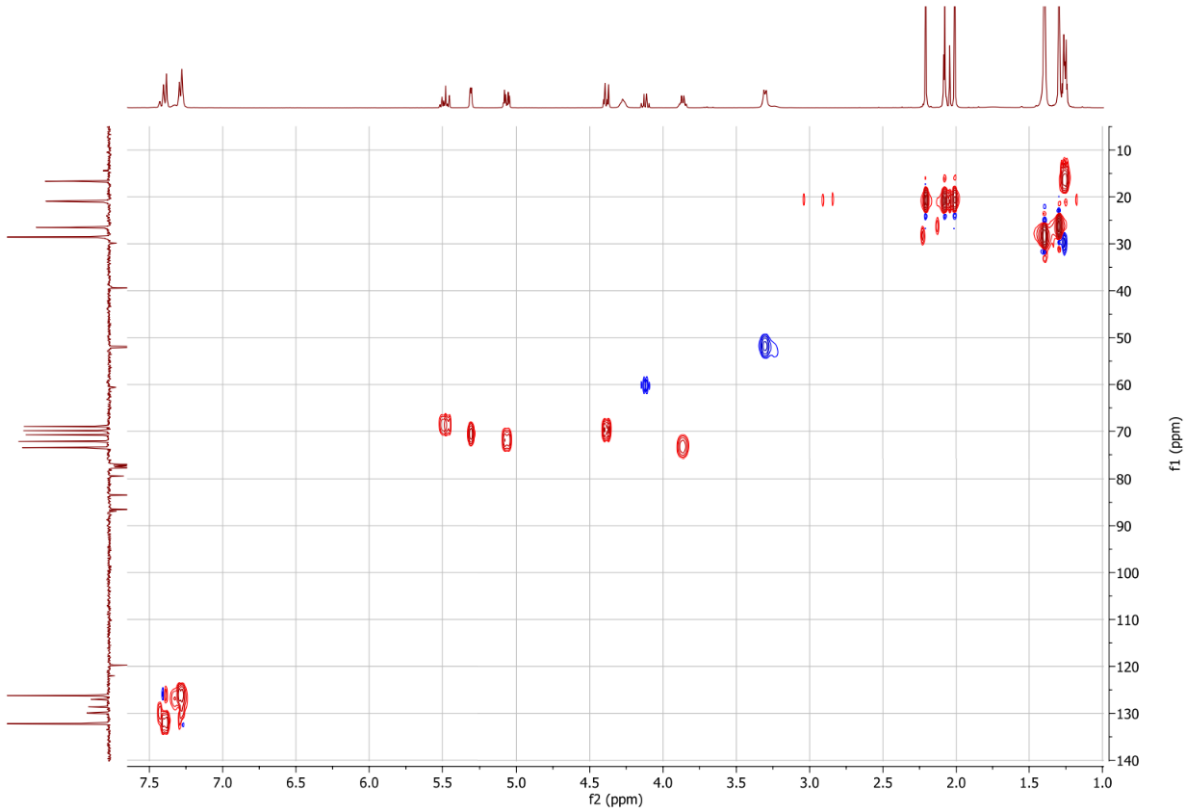


^{13}C NMR (400 MHz, $CDCl_3$), some shifts were extrapolated from the HSQC experiment:

δ = 170.9, 170.4, 169.7 (*C=O*), 156.2 (*C=O*), 148.1 (*C Ar*), 132.2 (*≡C-C-<u>CH</u> Ar*), 126.2 (*CH Ar*), 119.8 (*≡C-<u>C</u> Ar*), 86.5 (*C1-<u>C</u>≡*), 83.5 (*≡<u>C</u>-Ar*), 79.4 (*C tBu*), 73.5 (*C5*), 72.1 (*C3*), 70.7 (*C4*), 69.8 (*C1*), 68.9 (*C2*), 52.0 (CH_2), 39.4 (*<u>C</u>*(CH_3)$_2$), 28.5 (CH_3 *tBu*), 26.5 (2xCH_3), 21.0, 20.9, 20.9 (CH_3 *OAc*), 16.7 (*C6*).



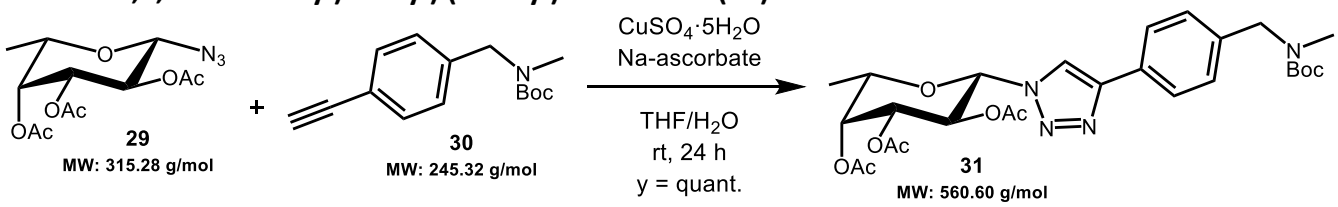
HSQC:



General procedure for copper-catalysed azide–alkyne cycloaddition (CuAAC):

H_2O was degassed with N_2 and used to prepare stock solutions of $CuSO_4{\cdot}5H_2O$ (0.04 M) and sodium ascorbate - Na-Asc (0.16 M) under N_2 atmosphere. THF was similarly degassed and used to dissolve the **acetylene-bearing compound** (1.0 eq, concentration: 0.2 M) under N_2 atmosphere. Equal parts of the two stock solutions were added to the THF solution, to match a THF/H_2O 1:1 ratio ($CuSO_4$: 0.1 eq, Na-Asc: 0.4 eq). Lastly, **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29** (1 eq) was added to the mixture and was stirred at room temperature overnight, until TLC showed completion. The reaction mixture was concentrated and re-dissolved in H_2O/DCM (1:1), then washed with a NH_3/NH_4Cl (1:1) solution to remove copper salts. The organic phase was dried over Na_2SO_4 and concentrated. The crude was used for the following step directly or after purification by flash or automatic chromatography (nHex/EtOAc).

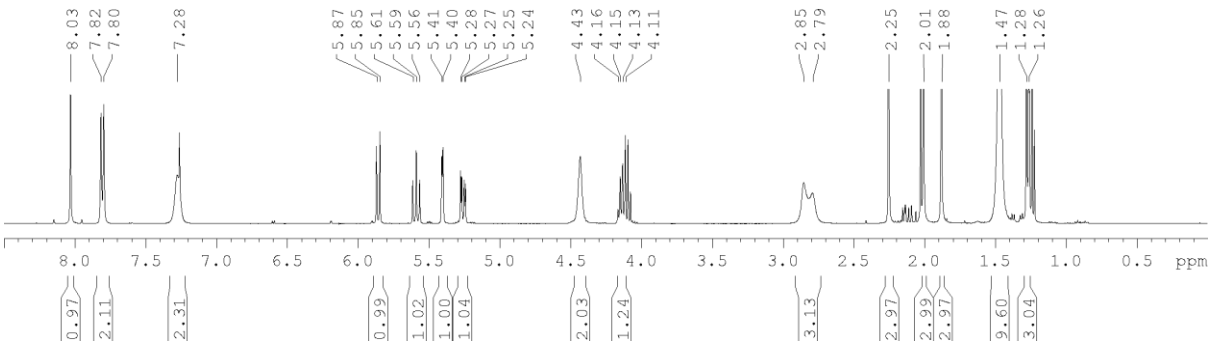
(31) Synthesis and characterization of ***tert*-butyl (4-(1-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)benzyl) (methyl) carbamate (31)**:



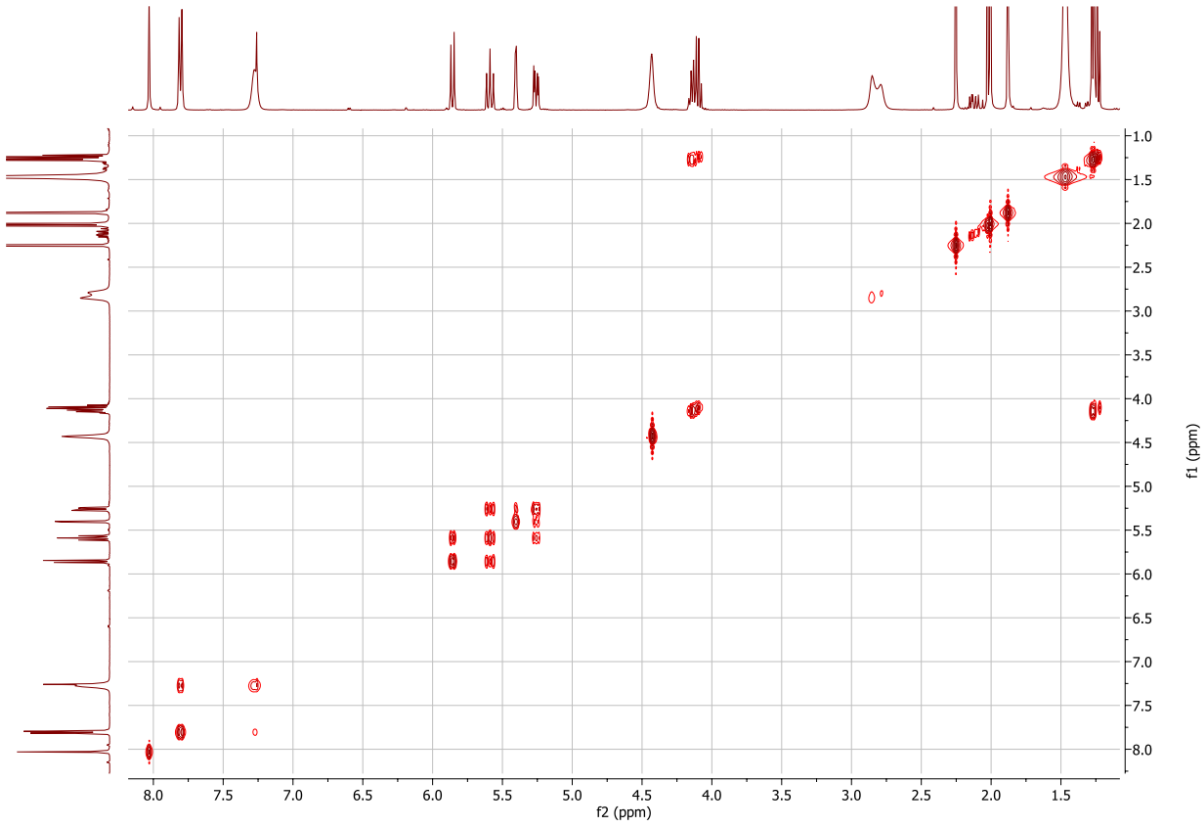
(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.14 mmol) was coupled to ***tert*-butyl (4-ethynylbenzyl)(methyl)carbamate 30** (0.14 mmol) following the protocol described, affording **31** (quantitative yield). TLC R_f (nHex/EtOAc: 6/4): 0.27. $[\alpha]_D^{19.6}$ = 51.13 ($CHCl_3$, 1). MS (ESI) calculated for $C_{27}H_{36}N_4O_9$ $[M + Na]^+$ *m/z*: 583.24; found: 583.26.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 8.03 (s, 1H, *CH hAr*), 7.81 (d, J = 8.0 Hz, 2H, *CH Ar*), 7.28 (bs, J = 8.1 Hz, 2H, *CH Ar*), 5.86 (d, $J_{1\text{-}2}$ = 9.3 Hz, 1H, *H-1*), 5.59 (t, $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.7 Hz, 1H, *H-2*), 5.40 (dd, $J_{4\text{-}3}$ = 3.2 Hz, 1H, *H-4*), 5.26 (dd, $J_{3\text{-}2}$ = 10.2 Hz, $J_{3\text{-}4}$ = 3.3 Hz, 1H, *H-3*), 4.43 (bs, 2H, CH_2), 3.99 (qd, $J_{5\text{-}CH3}$ = 6.4 Hz, 1H, *H-5*), 2.85 - 2.79 (bd, 3H, *N-CH_3*), 2.25, 2.01, 1.88 (s, 3H, *OAc*), 1.47 (bs, 9H, *tBu*), 1.27 (d, $J_{CH3\text{-}5}$ = 6.4 Hz, 3H, CH_3).

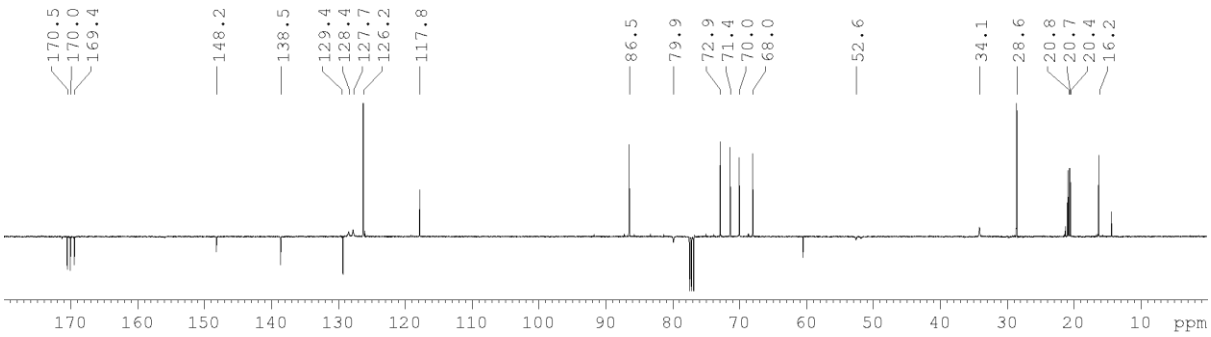


COSY:

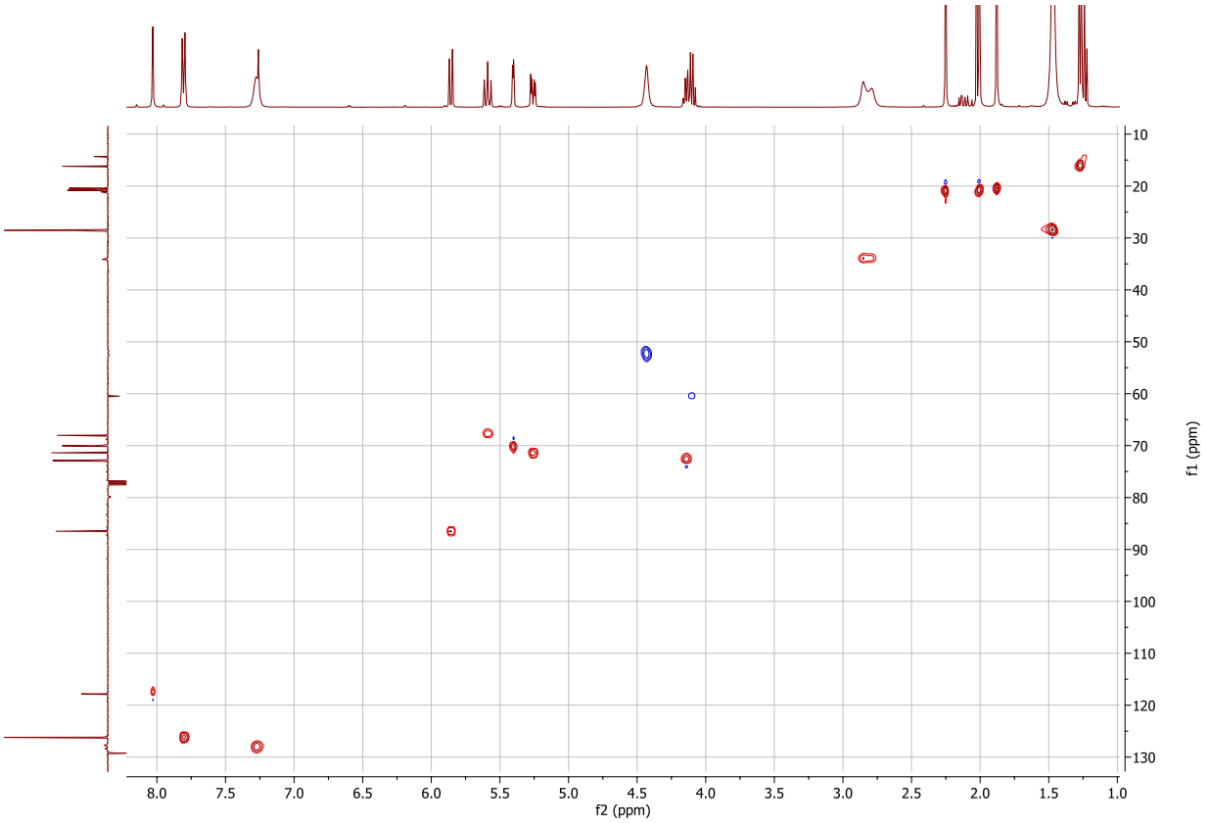


^{13}C NMR (400 MHz, $CDCl_3$):

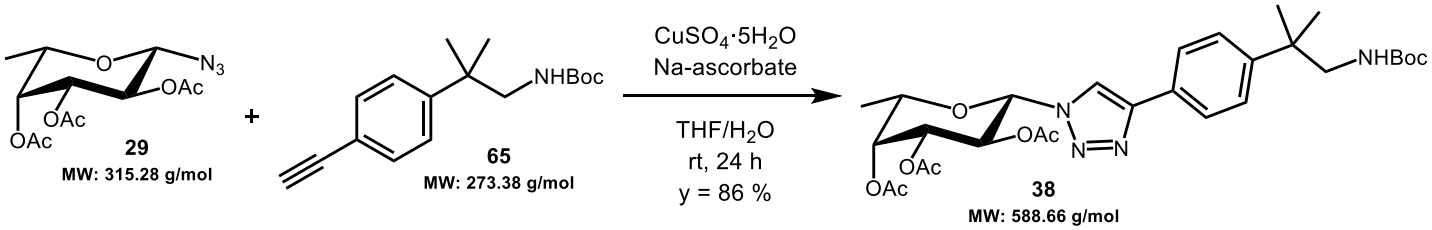
δ = 170.5, 170.0, 169.4 (*C=O*), 148.2 (*C hAr*), 138.6 (*CH_2-C Ar*), 129.2 (*hAr-C* Ar), 128.4 - 127.7 (*CH Ar*), 126.2 (*CH Ar*), 117.8 (*CH hAr*), 86.5 (*C1*), 79.9 (*C tBu*), 72.9 (*C5*), 71.4 (*C3*), 70.0 (*C4*), 68.0 (*C2*), 52.6 (*CH_2*), 34.1 (*CH_3*), 28.6 (*CH_3 tBu*), 20.8, 20.7, 20.4 (*CH_3 OAc*), 16.2 (*C6*).



HSQC:



(38) Synthesis and characterization of ***tert*-butyl (2-(4-(1-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)phenyl)-2-methylpropyl) carbamate (38)**:



(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.16 mmol) was coupled to ***tert*-butyl (2-(4-ethynylphenyl)-2-methylpropyl)carbamate 65** (0.16 mmol) following the protocol described, affording **38** (0.14 mmol, y = 86 %). TLC R_f (nHex/EtOAc: 6/4): 0.22.

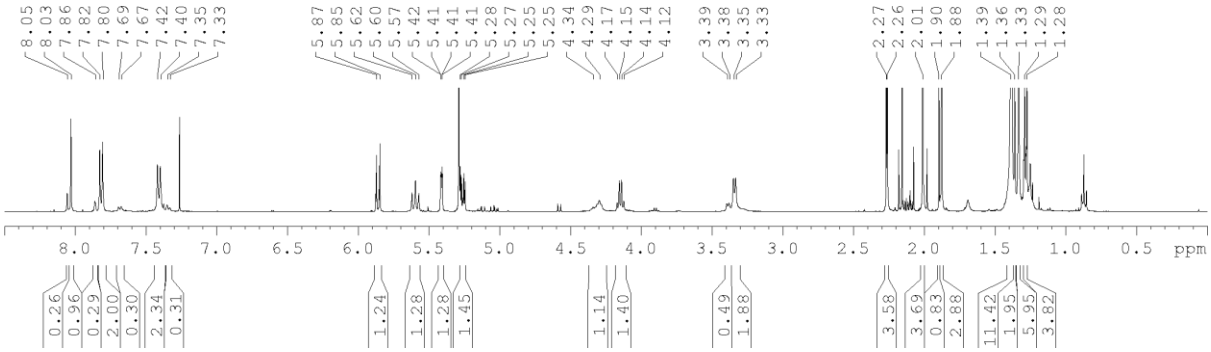
^{1}H NMR (400 MHz, $CDCl_3$): regioisomeric *para*/*meta* crude mixture (77/23)

para-isomer:

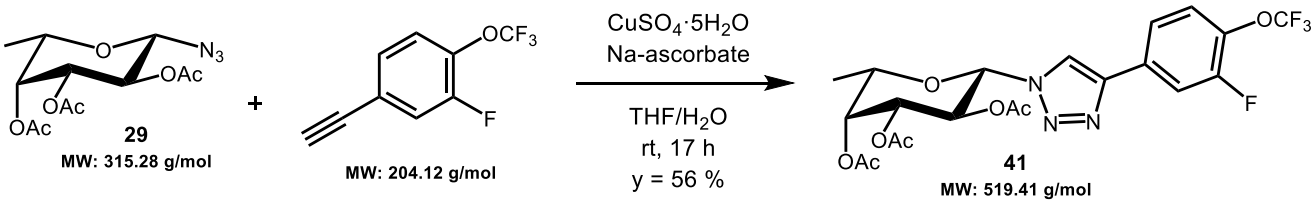
δ = 8.03 (s, 1H, *CH hAr*), 7.81 (d, J = 8.4 Hz, 2H, *CH Ar*), 7.41 (d, J = 8.4 Hz, 2H, *CH Ar*), 5.86 (d, $J_{1\text{-}2}$ = 9.4 Hz, 1H, *H-1*), 5.60 (dd, $J_{2\text{-}1}$ = 9.4 Hz, $J_{2\text{-}3}$ = 10.2 Hz, 1H, *H-2*), 5.41 (d, $J_{4\text{-}3}$ = 3.4 Hz, $J_{4\text{-}5}$ = 1.2 Hz, 1H, *H-4*), 5.26 (dd, $J_{3\text{-}2}$ = 10.3 Hz, $J_{3\text{-}4}$ = 3.4 Hz, 1H, *H-3*), 4.29 (bt, J = 6.1 Hz, 1H, *NH*), 4.15 (dq, $J_{5\text{-}4}$ = 1.2 Hz, $J_{5\text{-}CH3}$ = 6.4 Hz, 1H, *H-5*), 3.38 (bd, J = 6.3 Hz, 2H, *CH_2*), 2.26, 2.01, 1.88 (s, 3H, *OAc*), 1.39 (bs, 9H, *tBu*), 1.33 (s, 6H, 2x*CH_3*), 1.28 (d, $J_{6\text{-}5}$ = 6.4 Hz, 3H, *H-6*).

meta-isomer:

δ = 8.05 (s, 1H, *CH hAr*), 7.86 (bs, 1H, *CH Ar*), 7.68 (d, J = 7.6 Hz, 1H, *CH Ar*), 7.41 (m, 1H, *CH Ar*), 7.34 (m, 1H, *CH Ar*), 5.86 (d, $J_{1\text{-}2}$ = 9.3 Hz, 1H, *H-1*), 5.60 (t, 1H, *H-2*), 5.41 (d, 1H, *H-4*), 5.26 (dq, 1H, *H-3*), 4.34 (bt, 1H, *NH*), 4.14 (m, 1H, *H-5*), 3.38 (bd, 2H, CH_2), 2.27, 2.01, 1.90 (s, 3H, *OAc*), 1.39 (bs, 9H, *tBu*), 1.36 (s, 6H, 2xCH_3), 1.29 (d, 3H, *H-6*).



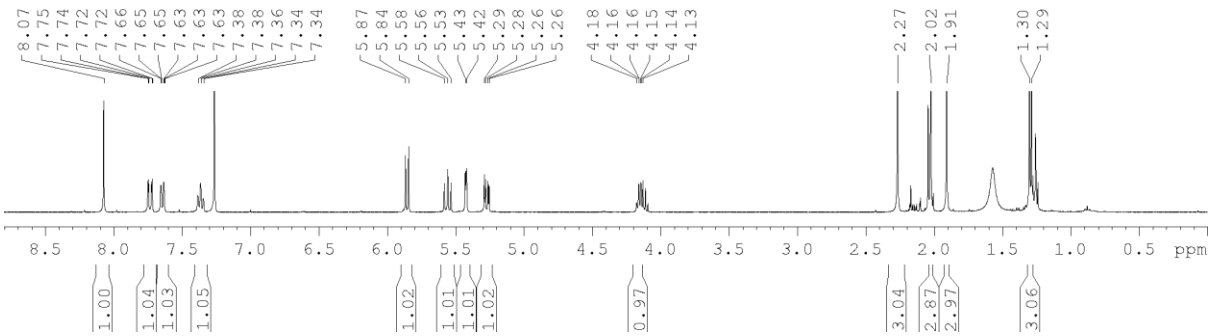
(41) Synthesis and characterization of **4-(3-fluoro-4-(trifluoromethoxy)phenyl)-1-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-1,2,3-triazole (41)**:



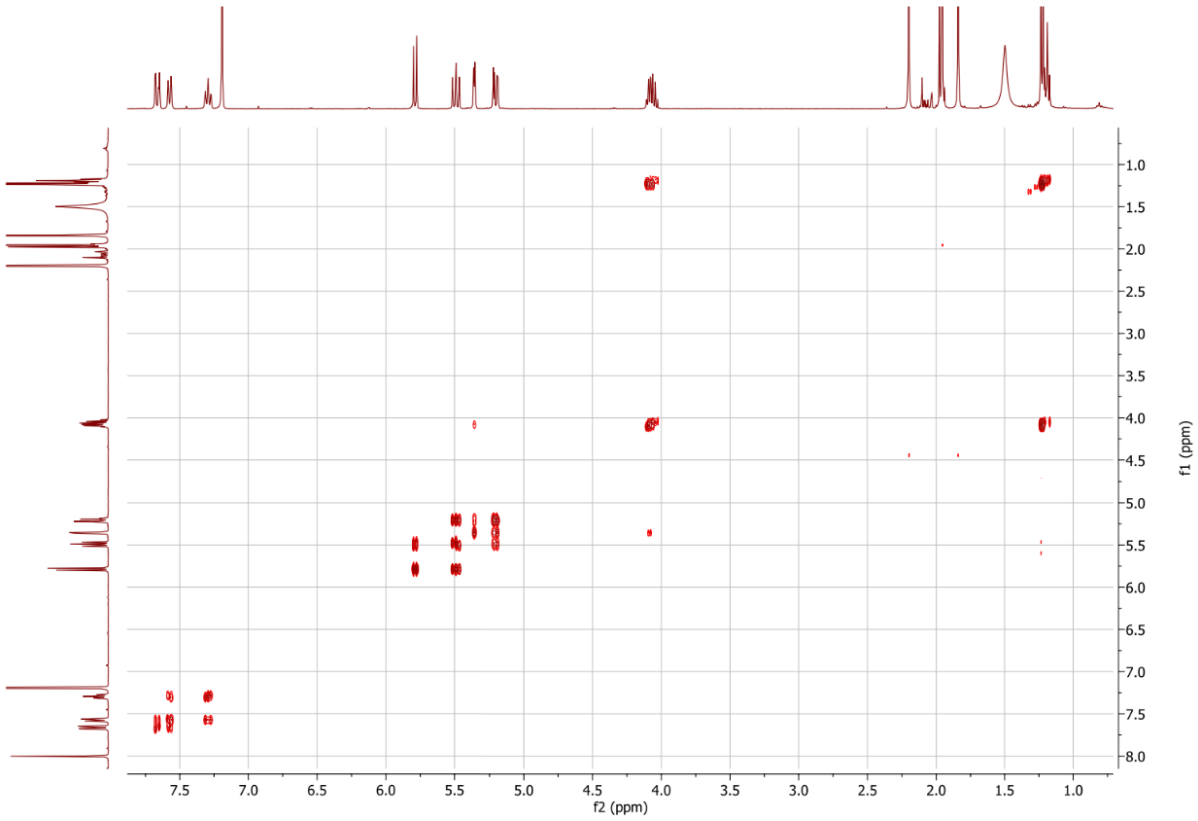
(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.16 mmol) was coupled to **4-ethynyl-2-fluoro-1-(trifluoromethoxy)benzene** (0.24 mmol) following the protocol described, affording **41** (0.09 mmol y = 56 %). TLC R_f (nHex/EtOAc: 7/3): 0.27.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 8.07 (s, 1H, *CH hAr*), 7.73 (d, J = 10.9 Hz, 1H, *CH Ar*), 7.64 (d, J' = 8.5 Hz, 1H, *CH Ar*), 7.36 (t, J' = 8.2 Hz, 1H, *CH Ar*), 5.86 (d, $J_{1\text{-}2}$ = 9.3 Hz, 1H, *H-1*), 5.56 (dd, $J_{2\text{-}1}$ = 9.3 Hz, $J_{2\text{-}3}$ = 10.3 Hz, 1H, *H-2*), 5.42 (d, $J_{4\text{-}3}$ = 3.4 Hz, $J_{4\text{-}5}$ = 1.2 Hz, 1H, *H-4*), 5.27 (dd, $J_{3\text{-}2}$ = 10.3 Hz, $J_{3\text{-}4}$ = 3.4 Hz, 1H, *H-3*), 4.15 (dq, $J_{5\text{-}4}$ = 1.2 Hz, $J_{5\text{-}CH3}$ = 6.4 Hz, 1H, *H-5*), 2.27, 2.02, 1.91 (s, 3H, *OAc*), 1.30 (2, 3H, $J_{CH3\text{-}5}$ = 6.4 Hz, CH_3).



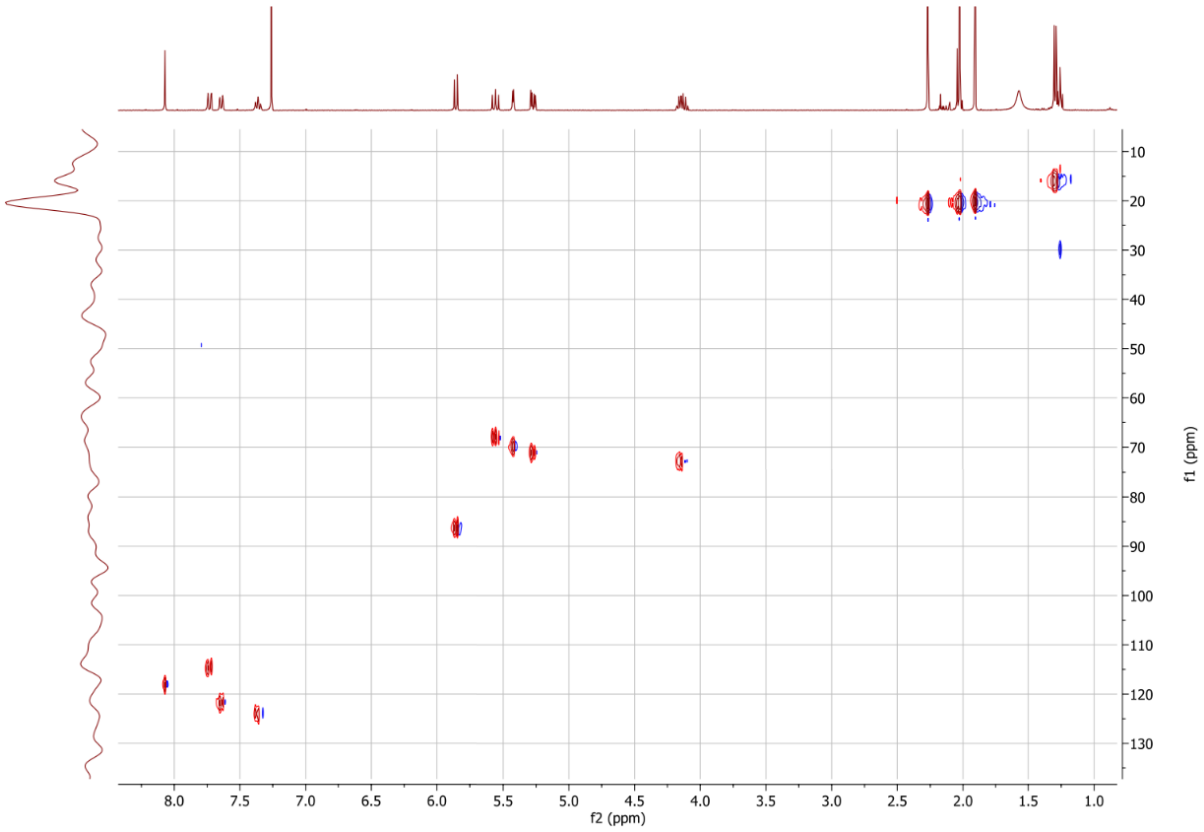
COSY:



^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 124.1, 121.7, 114,7 (*CH Ar*), 118.2 (*CH hAr*), 86.3 (*C1*), 72.9 (*C5*), 71.1 (*C3*), 69.9 (*C4*), 67.9 (*C2*), 20.5 (2xCH_3), 16.0 (*C6*).

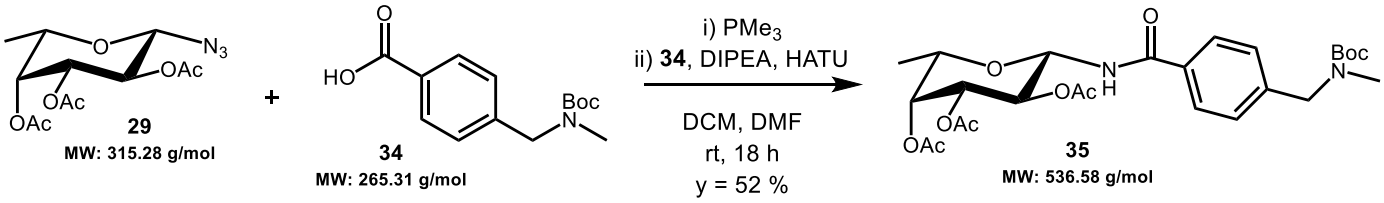
HSQC:



General procedure for amide bond formation through Staudinger ligation, adapted from Bianchi and co-workers:[232]

To the **(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29** (1.1 eq) in dry DCM (concentration: 0.15 M) was added a 1 M solution of PMe_3 in toluene (1.5 eq) under N_2 atmosphere. The reaction mixture stirred at room temperature for 0.5 - 1 h, until TLC showed full reduction, before being concentrated. On a second flask, the **carboxylic fragment** (1.0 eq) was dissolved in dry DMF (concentration: 0.15 M) under N_2 atmosphere. N,N-Diisopropylethylamine - DIPEA (2.2 eq) and Hexafluorophosphate Azabenzotriazole Tetramethyl Uronium - HATU (1.1 eq) were added to the second flask and stirred for 1 h before being added to the reduced fucoside, adding DMF (concentration: 0.10 M). The resulting solution was stirred overnight, then was concentrated, redissolved in DCM, and washed with a 1 M aqueous HCl solution, then a $NaHCO_3$ saturated aqueous solution. The organic phase was dried over Na_2SO_4 and purified by automatic chromatography (nHex/EtOAc).

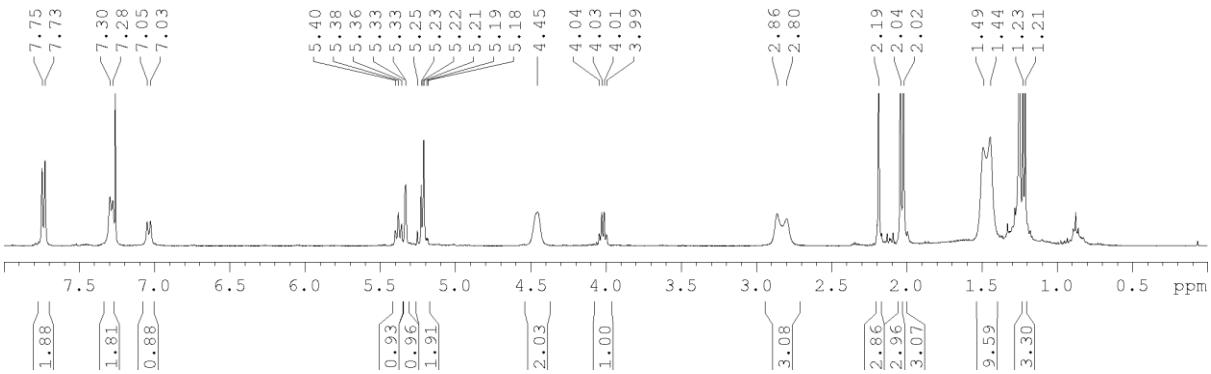
(35) Synthesis and characterization of ***tert*-butyl (4-((2,3,4-tri-O-acetyl β-L-fucopyranosyl)carbamoyl)benzyl) (methyl) carbamate (35)**:



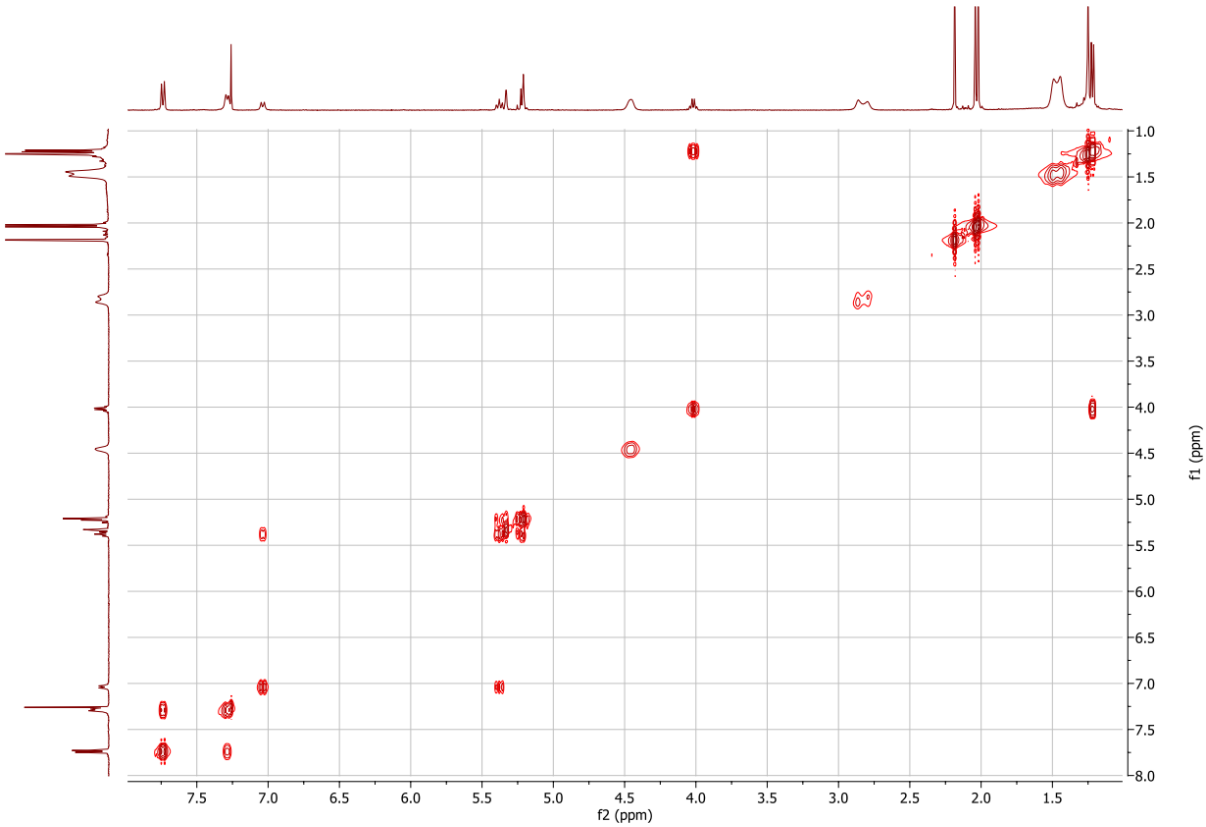
(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.154 mmol) was coupled to **4-(Boc-aminomethyl)benzoic acid 34** (0.139 mmol) following the protocol described, affording **35** (0.072 mmol, y = 52 %). TLC R_f (nHex/EtOAc: 6/4): 0.10. $[\alpha]_D^{17}$ = 7.23 ($CHCl_3$, c 1). MS (ESI) calculated for $C_{26}H_{36}N_2O_{10}$ [M + Na]+ m/z: 559.23; found: 559.30.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.74 (d, J = 8.2 Hz, 2H, *CH Ar*), 7.29 (d, J = 7.6 Hz, 2H, *CH Ar*), 7.04 (d, J_{NH-1} = 8.9 Hz, 1H, *NH*), 5.38 (d, J_{1-2} = 9.0 Hz, 1H, *H-1*), 5.33 (d, J_{4-3} =2.4 Hz, 2H, *H-4*), 5.25 - 5.18 (mult., J_{2-1} = J_{2-3} = 10.3 Hz, 1H, *H-2* + *H-3*), 4.45 (bs, 2H, CH_2), 4.02 (q, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 2.86 - 2.80 (bd, 3H, *N-CH_3*), 2.19, 2.04, 2.02 (s, 3H, *OAc*), 1.49 - 1.44 (bs, 9H, *tBu*), 1.22 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*).

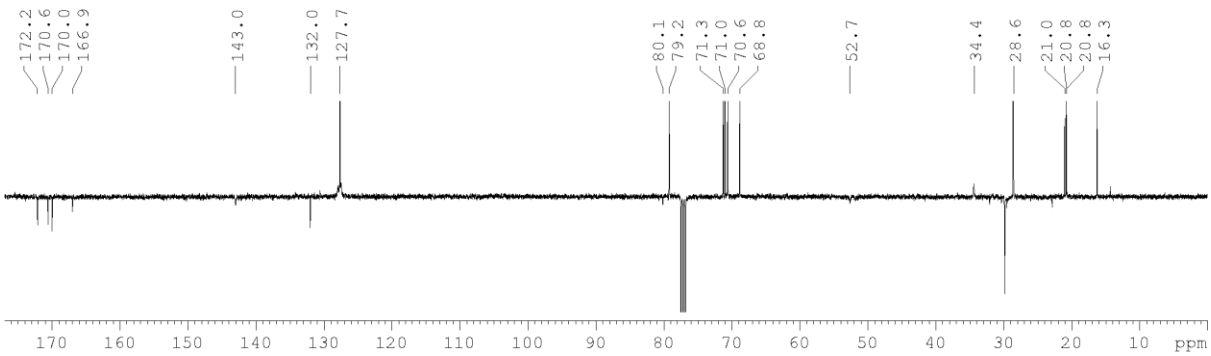


COSY:

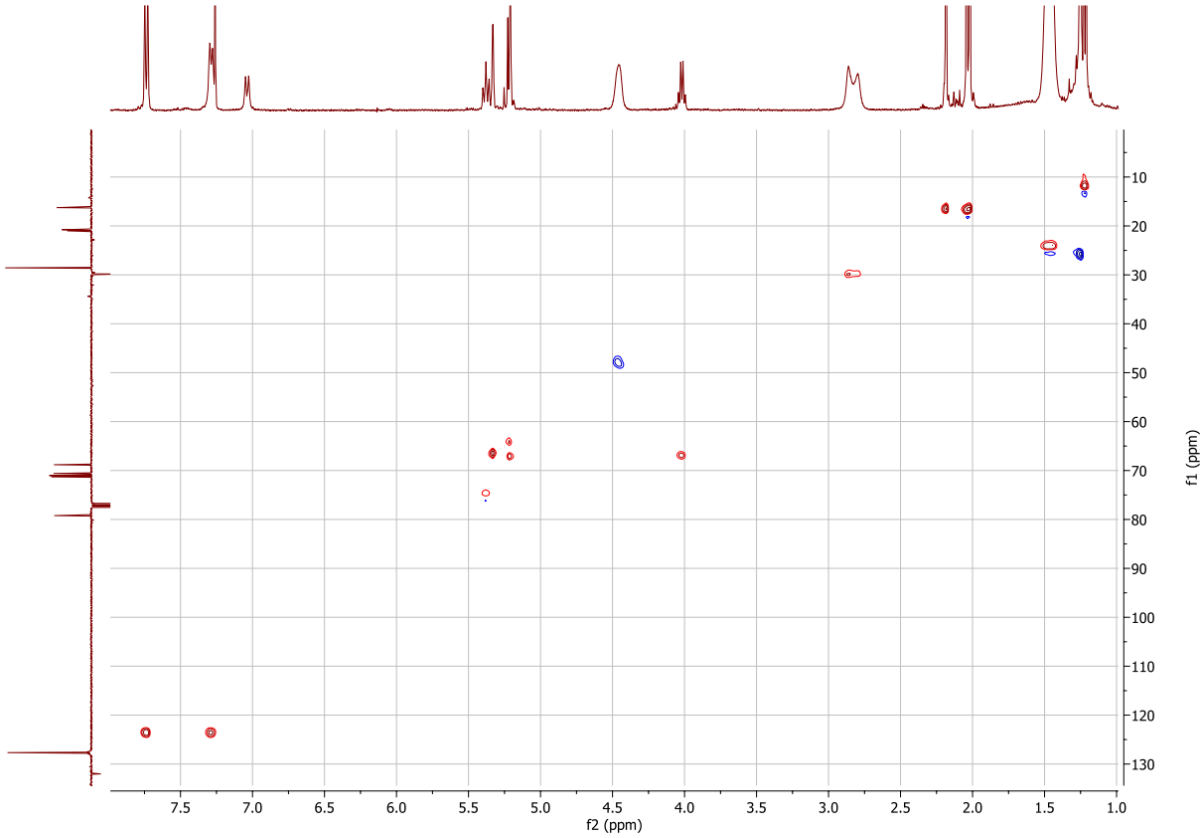


^{13}C NMR (400 MHz, $CDCl_3$):

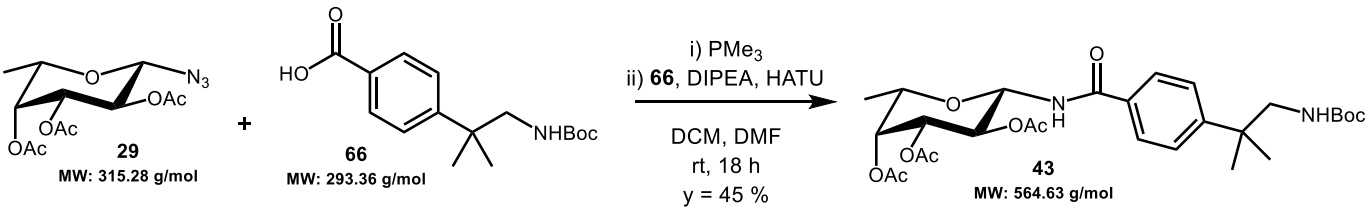
δ = 172.2, 170.6, 170.0 (C=O Ac), 166.9 (NH-C=O), 143.0 (*CH_2-C Ar*), 132.0 (O=*C-C Ar*), 127.7 (*CH Ar*), 80.1 (*C tBu*), 79.2 (*C1*), 71.3 (*C3*), 71.0 (*C5*), 70.6 (*C4*), 68.8 (*C2*), 52.7 (*CH_2*), 34.4 (*CH_3*), 28.6 (*CH_3 tBu*), 21.0, 20.8, 20.8 (*CH_3 OAc*), 16.3 (*C6*).



HSQC:



(43) Synthesis and characterization of ***tert*-butyl (2-(4-((2,3,4-tri-O-acetyl β-L-fucopyranosyl)carbamoyl)phenyl)-2-methylpropyl) carbamate (43)**:



(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.20 mmol) was coupled to **4-(1-((*tert*-butoxycarbonyl)amino)-2-methylpropan-2-yl)benzoic acid 66** (0.17 mmol) following the protocol described, affording **43** (0.08 mmol, y = 45 %). TLC R_f (nHex/EtOAc: 6/4): 0.10. MS (ESI) calculated for $C_{28}H_{40}N_2O_{10}$ $[M + Na]^+$ *m/z*: 587.26; found: 587.34.

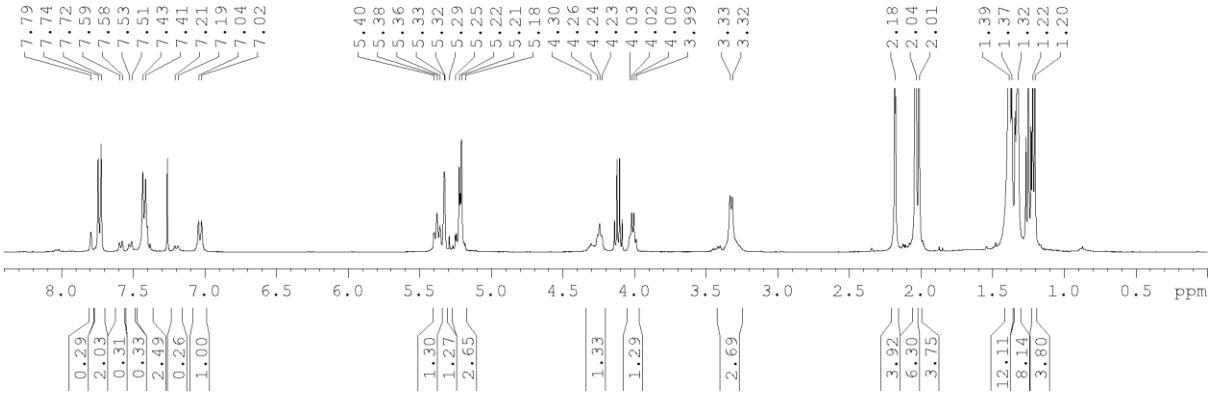
^{1}H NMR (400 MHz, $CDCl_3$): regioisomeric *para*/*meta* mixture (77/23)

para-isomer:

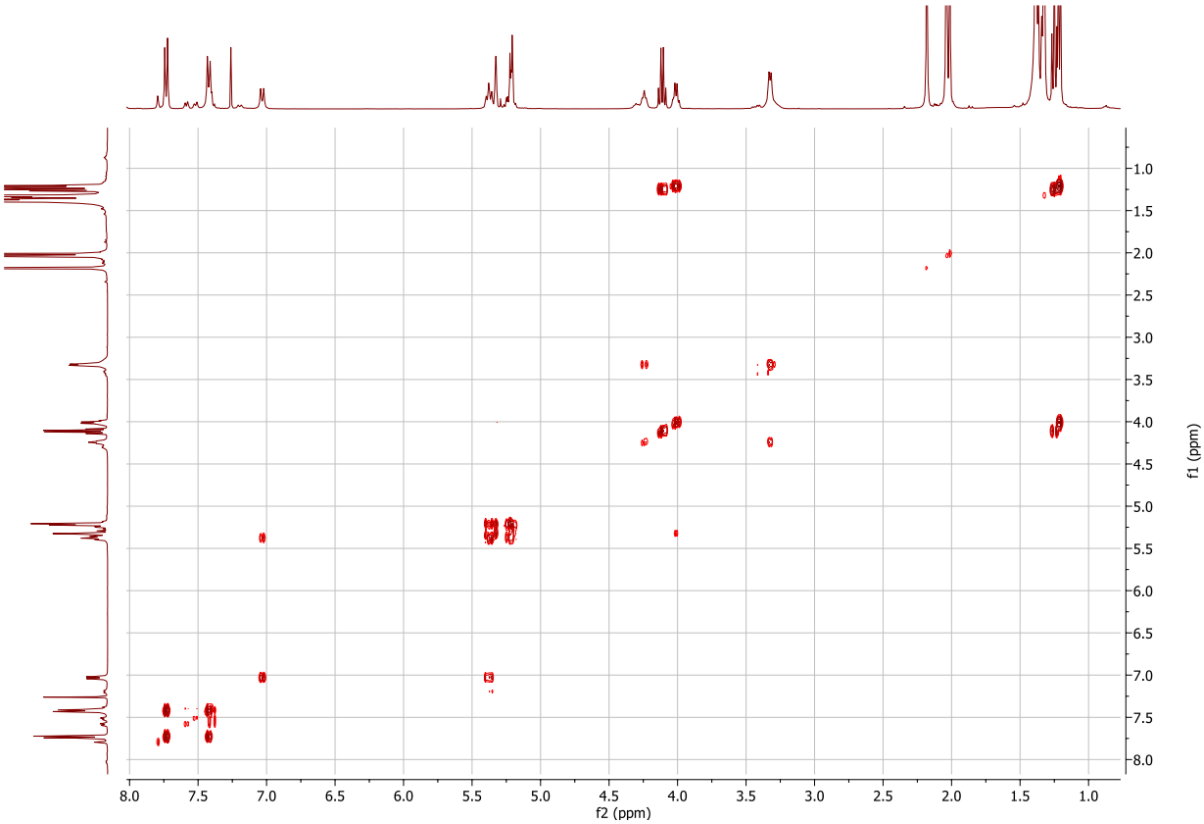
δ = 7.73 (d, J = 8.5 Hz, 2H, *CH Ar*), 7.42 (d, J = 8.4 Hz, 2H, *CH Ar*), 7.03 (d, J_{NH-1} = 9.0 Hz, 1H, *NH*), 5.38 (t, J_{1-NH} = J_{1-2} = 9.0 Hz, 1H, *H-1*), 5.32 (dd, J_{4-3} = 2.9 Hz J_{4-5} = 1.2 Hz, , 1H, *H-4*), 5.27 - 5.18 (mult., 1H, *H-2* + *H-3*), 4.24 (t, J_{NH-CH2} = 6.4 Hz, 1H, *NH-Boc*), 4.01 (dq, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.4 Hz, 1H, *H-5*), 3.31 (bd, J_{NH-CH2} = 6.4 Hz, 2H, CH_2), 2.18, 2.04, 2.01 (s, 3H, *OAc*), 1.39 (bs, 9H, *tBu*), 1.32 (bs, 6H, 2xCH_3), 1.26 (d, J_{CH3-5} = 6.4 Hz, 3H, *H-6*).

meta-isomer:

δ = 7.97 (bs, 1H, *CH Ar*), 7.59 (d, J = 7.6 Hz, 1H, *Ar*), 7.52 (d, J = 7.6 Hz, 1H, *Ar*), 7.41 (m, H, *Ar*), 7.20 (d, J_{NH-1} = 8.8 Hz, 1H, *NH*), 5.38 (t, 1H, *H-1*), 5.32 (d, 1H, *H-4*), 5.27 - 5.18 (mult., 1H, *H-2* + *H-3*), 4.30 (t, 1H, *NH-Boc*), 4.01 (dq, 1H, *H-5*), 3.31 (bd, 2H, CH_2), 2.18, 2.04, 2.01 (s, 3H, *OAc*), 1.39 (bs, 9H, *tBu*), 1.32 (bs, 6H, 2xCH_3), 1.21 (d, 3H, *H-6*).



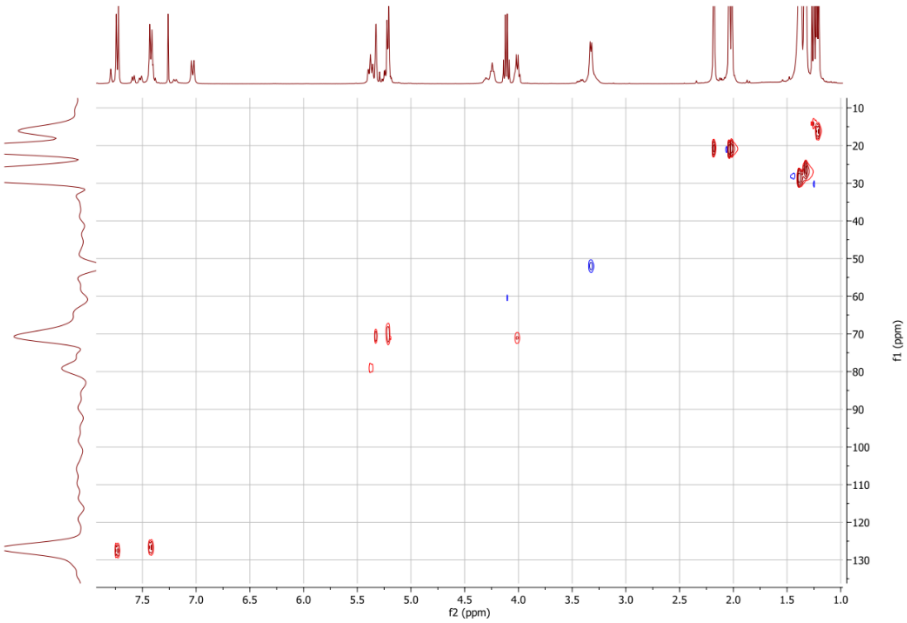
COSY:



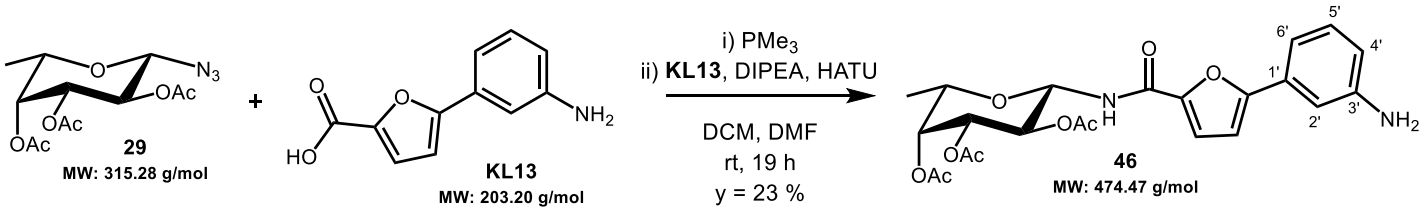
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 127.5, 126.6 (*para CH Ar*), 130.2, 125.3, 125.0, 126.6 (*meta CH Ar*), 79.0 (*C1*), 71.1 (*C5*), 70.6 (*C4*), 71.0 - 68.8 (*C3*,*C2*), 52.0 (CH_2), 28.6 (CH_3 *tBu*), 26.5 (2xCH_3), 20.9 (CH_3 *OAc*), 16.2 (*C6*).

HSQC:



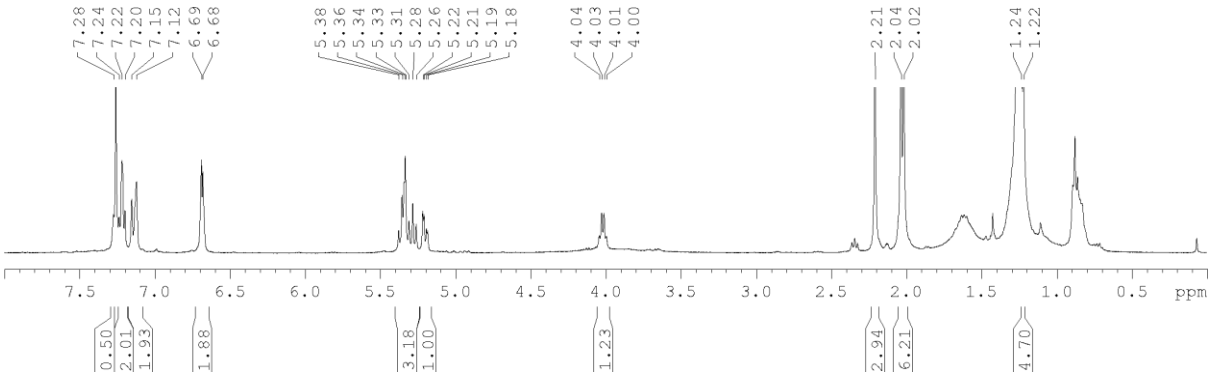
(46) Synthesis and characterization of **5-(3-aminophenyl)furan-2-carboxamido-(2,3,4-tri-O-acetyl-β-L-fucopyranose) (46)**:



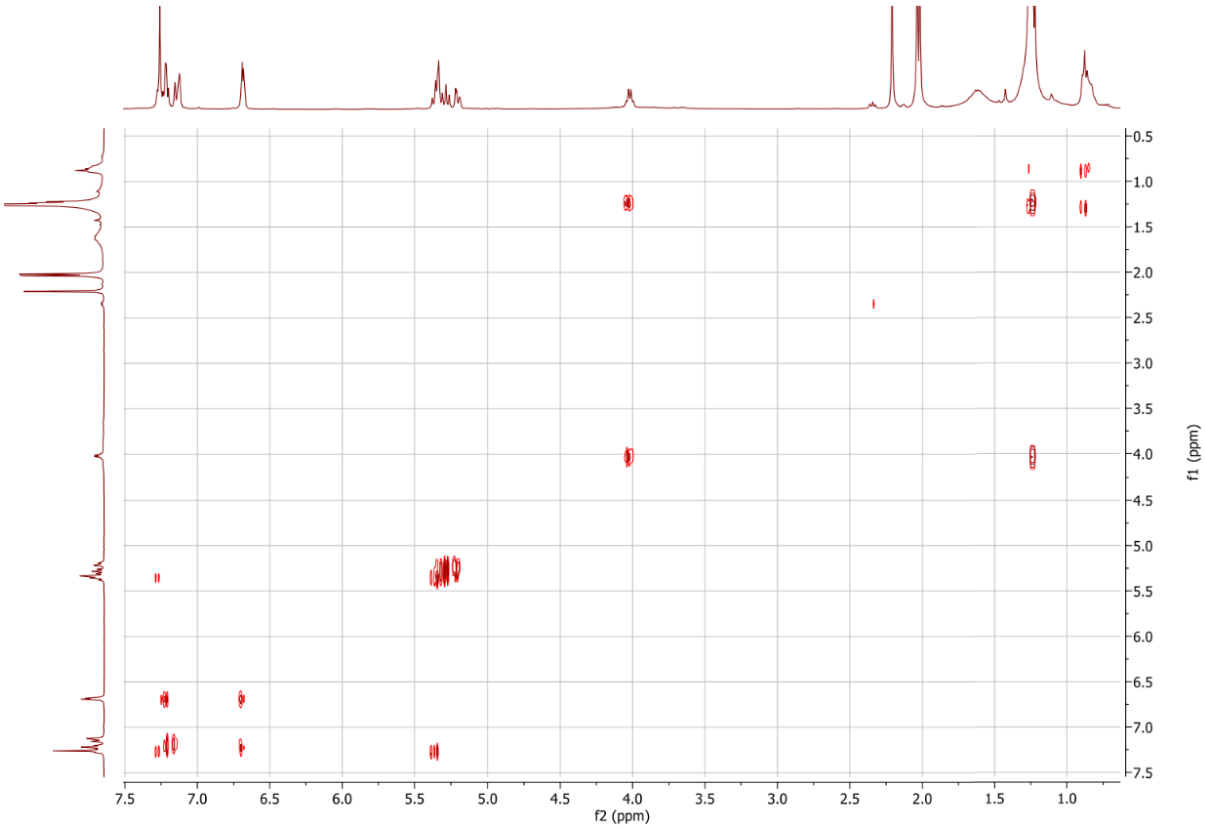
(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.082 mmol) was coupled to **5-(3-aminophenyl)furan-2-carboxylic acid KL12** (0.064 mmol) following the protocol described, affording **46** (0.015 mmol, y = 23 %). TLC R_f (nHex/EtOAc: 6/4): 0.35. MS (ESI) calculated for $C_{23}H_{26}N_2O_9$ $[M + H]^+$ *m/z*: 475.17; found: 474.75. $[M - H]^-$ *m/z*: 473.16; found: 472.85.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.28 (d, 1H, *NH*), 7.22 (mult., J_{ortho} = 7.9 Hz, 2H, *CH Ar*), 7.15 - 7.12 (mult., 2H, *CH Ar*), 6.69 (mult., J = 3.8 Hz, 2H, *CH hAr*), 5.38 - 5.26 (mult., J_{2-1} = 10.0 Hz, 3H, *H-1* + *H-2* + *H-4*), 5.20 (dd, J_{3-2} = 9.8 Hz, J_{3-4} = 3.1 Hz, 1H, *H-3*), 4.02 (dq, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 2.21, 2.04, 2.02 (s, 3H, *OAc*), 1.23 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*).



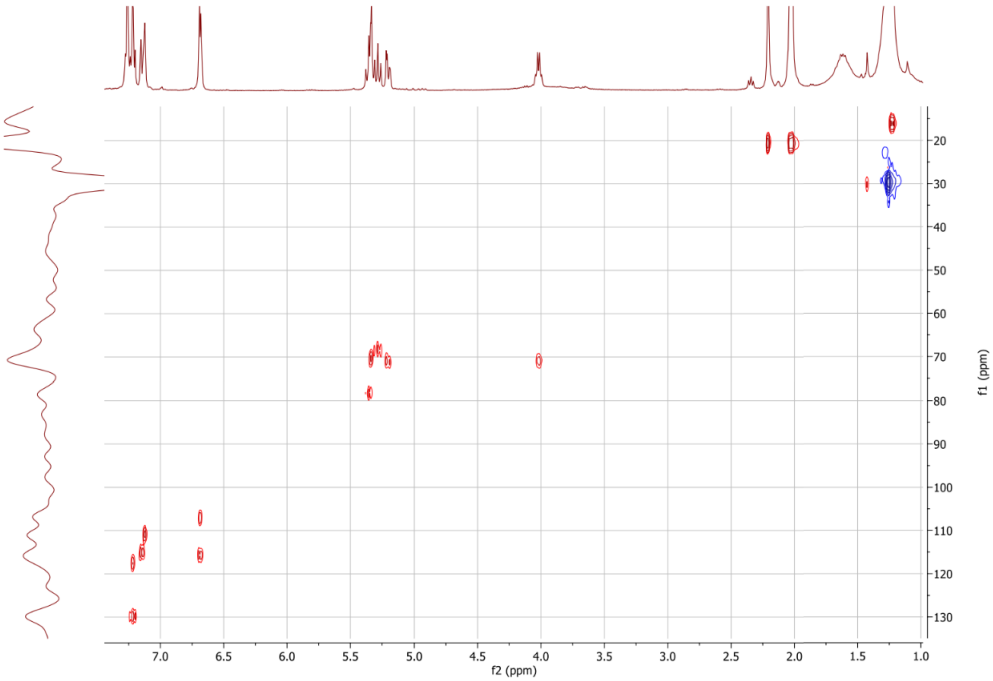
COSY:



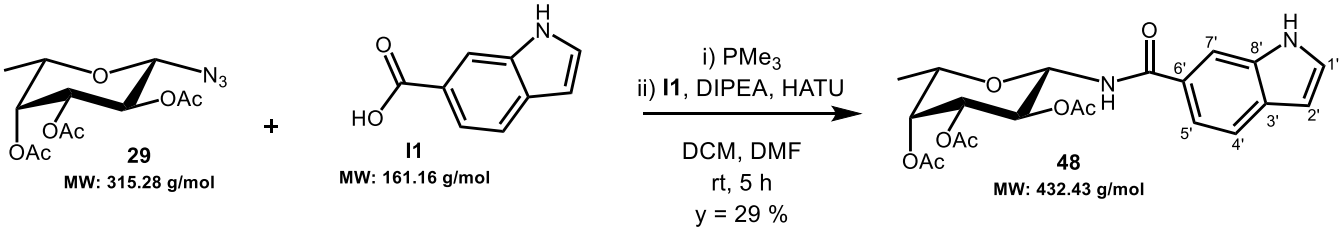
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 129.8 (*C5'*), 117.6 (*C4'*), 115.6 (CH_{furane}), 115.1 (*C6'*), 110.9 (*C2'*), 107.0 (CH_{furane}), 78.3 (*C1*), 71.0 (*C3*), 70.9 (*C5*), 70.4 (*C4*), 68.3 (*C2*), 20.7 (CH_3 *OAc*), 16.0 (*C6*)

HSQC:



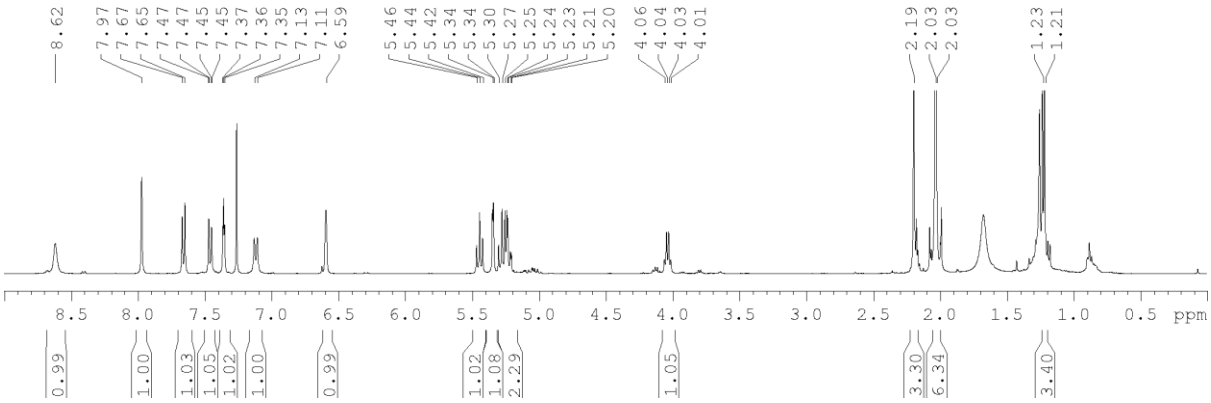
(48) Synthesis and characterization of **N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-indole-6-carboxamide (48)**:



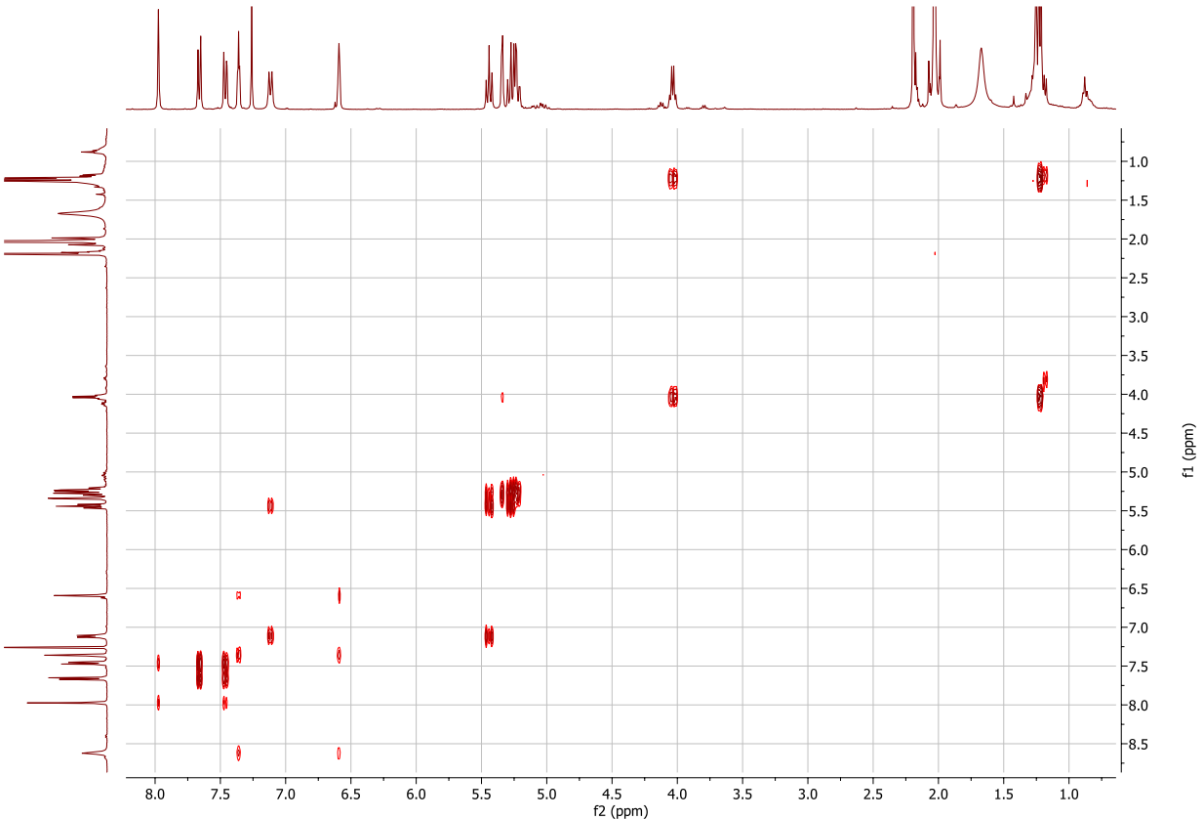
(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.079 mmol) was coupled to **1H-indole-6-carboxylic acid I1** (0.095 mmol) following the protocol described, affording **48** (0.023 mmol, y = 29 %). TLC R_f (nHex/EtOAc: 1/1): 0.27.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 8.62 (bs, 1H, *NH Ar*), 7.97 (s, 1H, *H-4'*), 7.66 (d, J = 8.4 Hz, 1H, *H-5'*), 7.46 (dd, J = 8.4 Hz, J' = 1.1 Hz, 1H, *H-7'*), 7.36 (t, J'' = 2.7 Hz, 1H, *H-1'*), 7.12 (d, J_{NH-1} = 9.0 Hz, 1H, *NH-C=O*), 6.59 (m, 1H, *H-2'*), 5.44 (t, $J_{NH-1} = J_{1-2}$ = 8.9 Hz, 1H, *H-1*), 5.34 (d, J_{4-3} = 2.8 Hz, 1H, *H-4*), 5.30 - 5.20 (mult., 2H, *H-2* + *H-3*), 4.04 (dq, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 2.19, 2.03, 2.03 (s, 3H, *OAc*), 1.22 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).



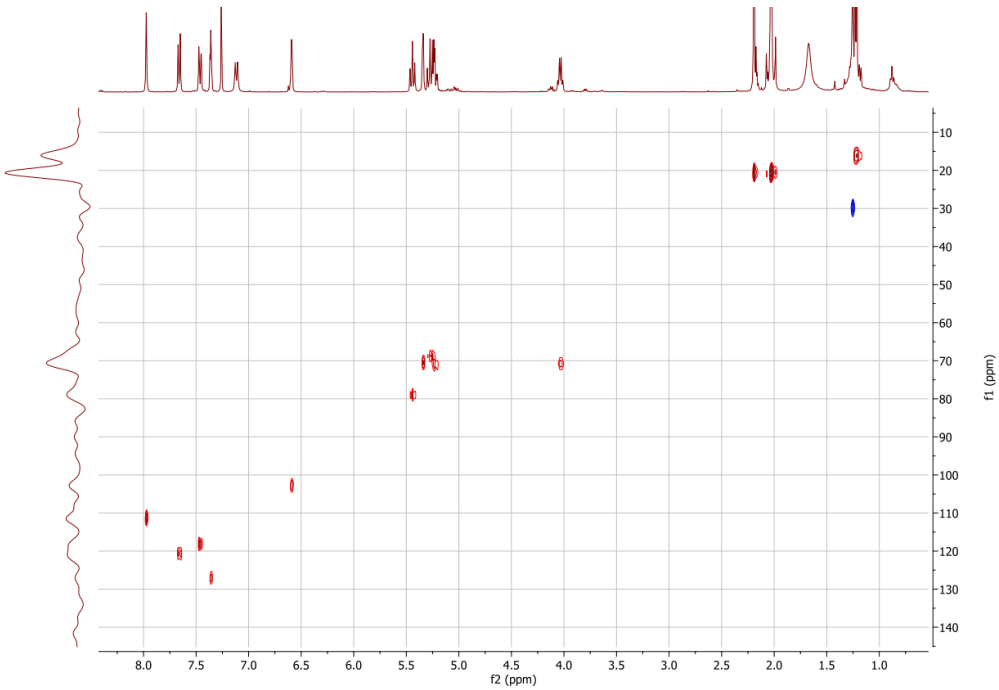
COSY:



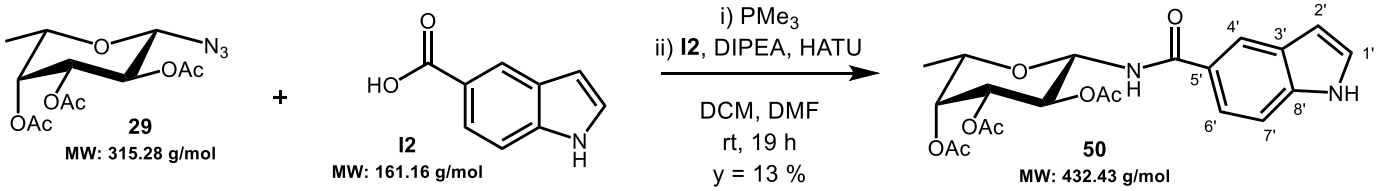
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 127.0 (*C1'*), 120.6 (*C5'*), 118.2 (*C7'*), 111.3 (*C4'*), 102.7 (*C2'*), 79.0 (*C1*), 71.1 (*C3*), 70.8 (*C5*), 70.5 (*C4*), 68.7 (*C2*), 20.5 (*CH_3 OAc*), 16.2 (*C6*).

HSQC:



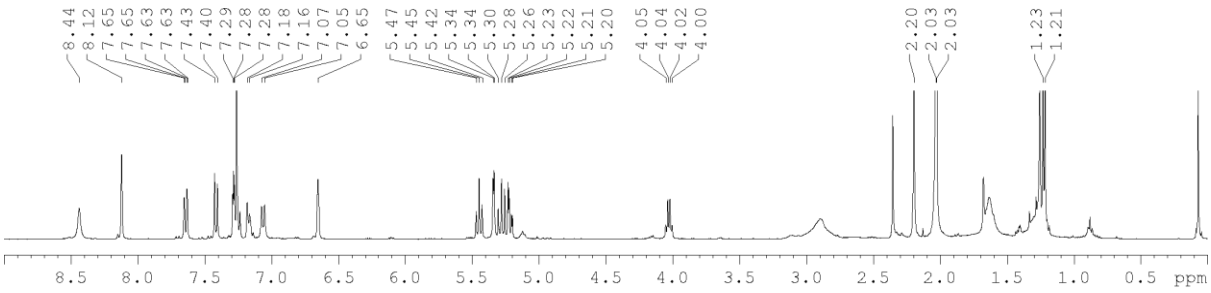
(50) Synthesis and characterization of **N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-indole-5-carboxamide (50)**:



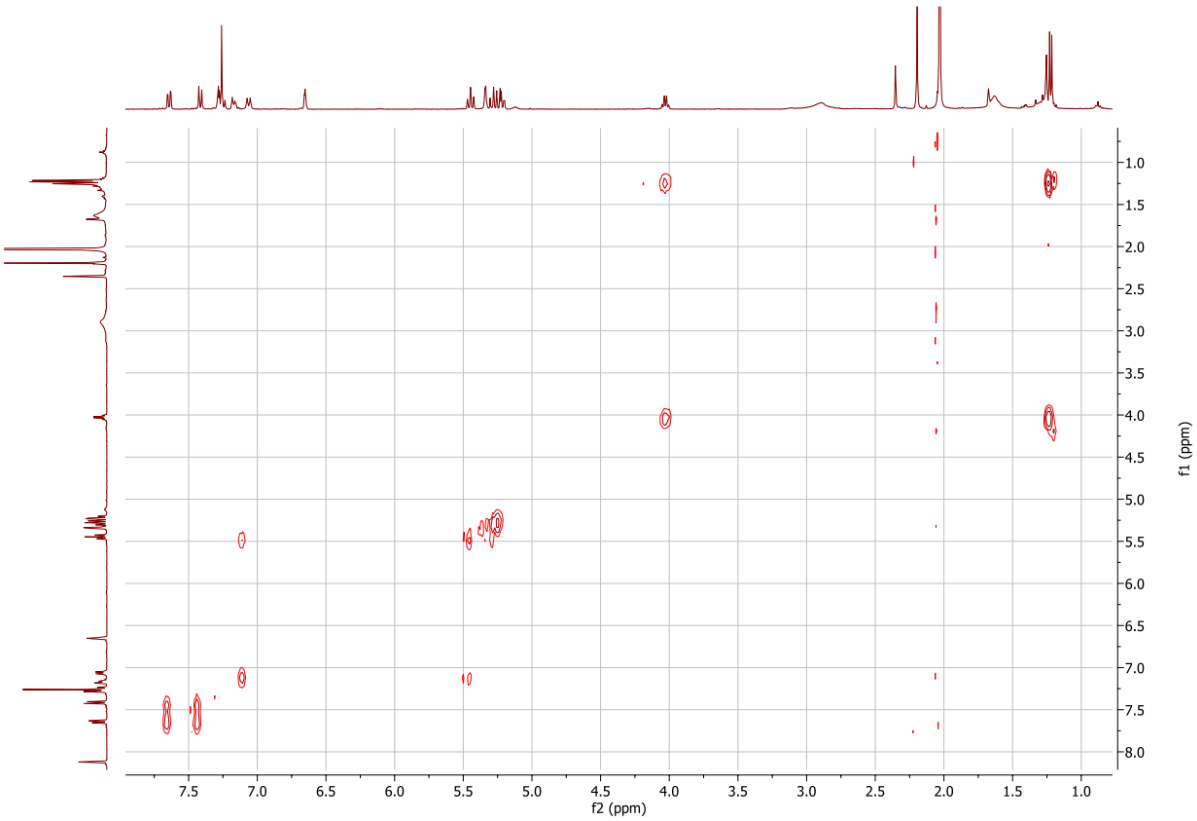
(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.104 mmol) was coupled to **1H-indole-5-carboxylic acid I2** (0.124 mmol) following the protocol described, affording **50** (0.014 mmol, y = 13 %). TLC R_f (nHex/EtOAc: 1/1): 0.16.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 8.44 (bs, 1H, *NH Ar*), 8.12 (s, 1H, *H-4'*), 7.64 (dd, J = 8.7 Hz, J' = 1.5 Hz, 1H, *H-6'*), 7.41 (d, J = 8.6 Hz, 1H, *H-7'*), 7.28 (t, J'' = 2.8 Hz, 1H, *H-1'*), 7.06 (d, J_{NH-1} = 9.0 Hz, 1H, *NH-C=O*), 6.65 (m, 1H, *H-2'*), 5.45 (t, J_{NH-1} = J_{1-2} = 9.0 Hz, 1H, *H-1*), 5.34 (d, J_{4-3} = 2.6 Hz, 1H, *H-4*), 5.28 (t, J_{2-1} = J_{2-3} = 9.0 Hz, 1H, *H-2*), 5.21 (dd, J_{3-2} = 10.3 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 4.03 (dq, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 2.20, 2.03, 2.03 (s, 3H, *OAc*), 1.22 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*).



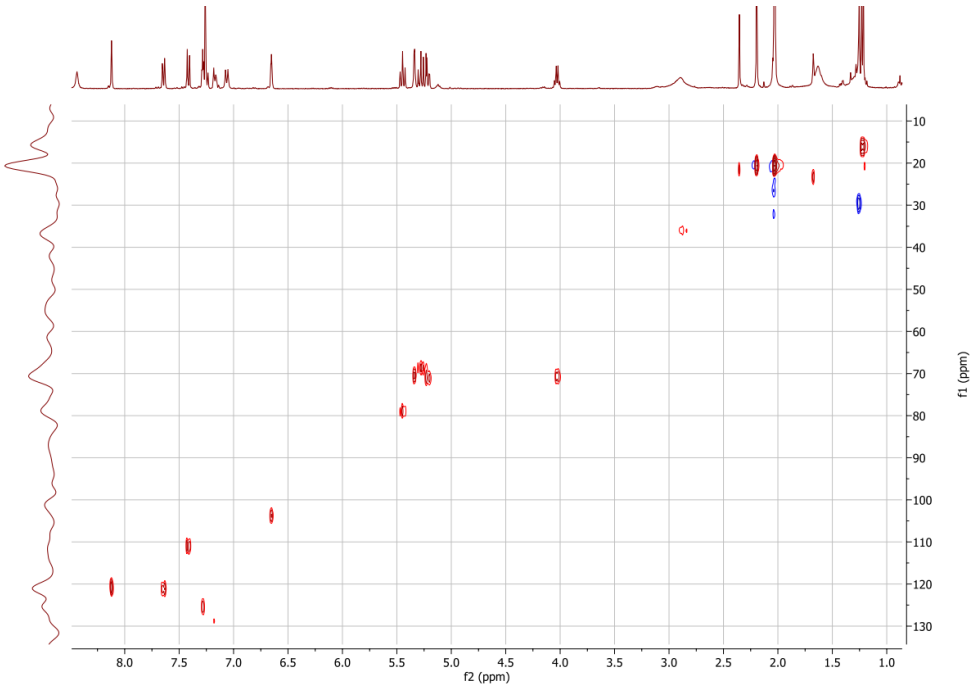
COSY:



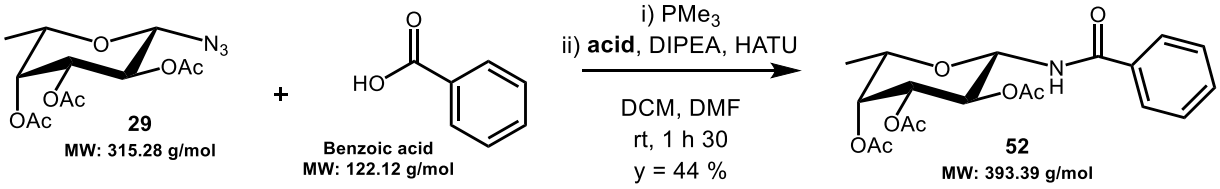
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 125.3 (*C1'*), 121.2 (*C6'*), 120.8 (*C4'*), 111.0 (*C7'*), 103.8 (*C2'*), 79.1 (*C1*), 71.0 (*C3*), 70.8 (*C5*), 70.4 (*C4*), 68.6 (*C2*), 20.6 (*CH_3 OAc*), 16.1 (*C6*).

HSQC:



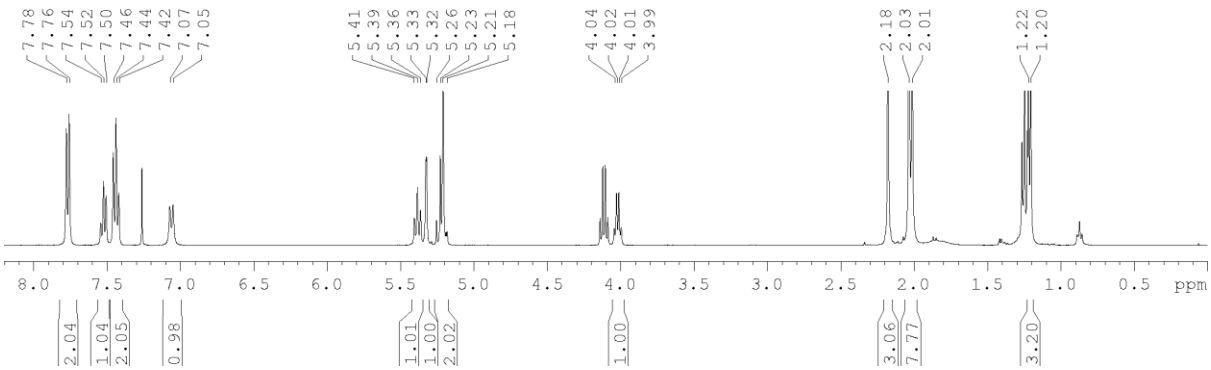
(52) Synthesis and characterization of **N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-benzamide (52)**:



(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.070 mmol) was coupled to **benzoic acid** (0.105 mmol) following the protocol described, affording **52** (0.031 mmol, y = 44 %). TLC R_f (nHex/EtOAc: 6/4): 0.23.

^{1}H NMR (400 MHz, $CDCl_3$):

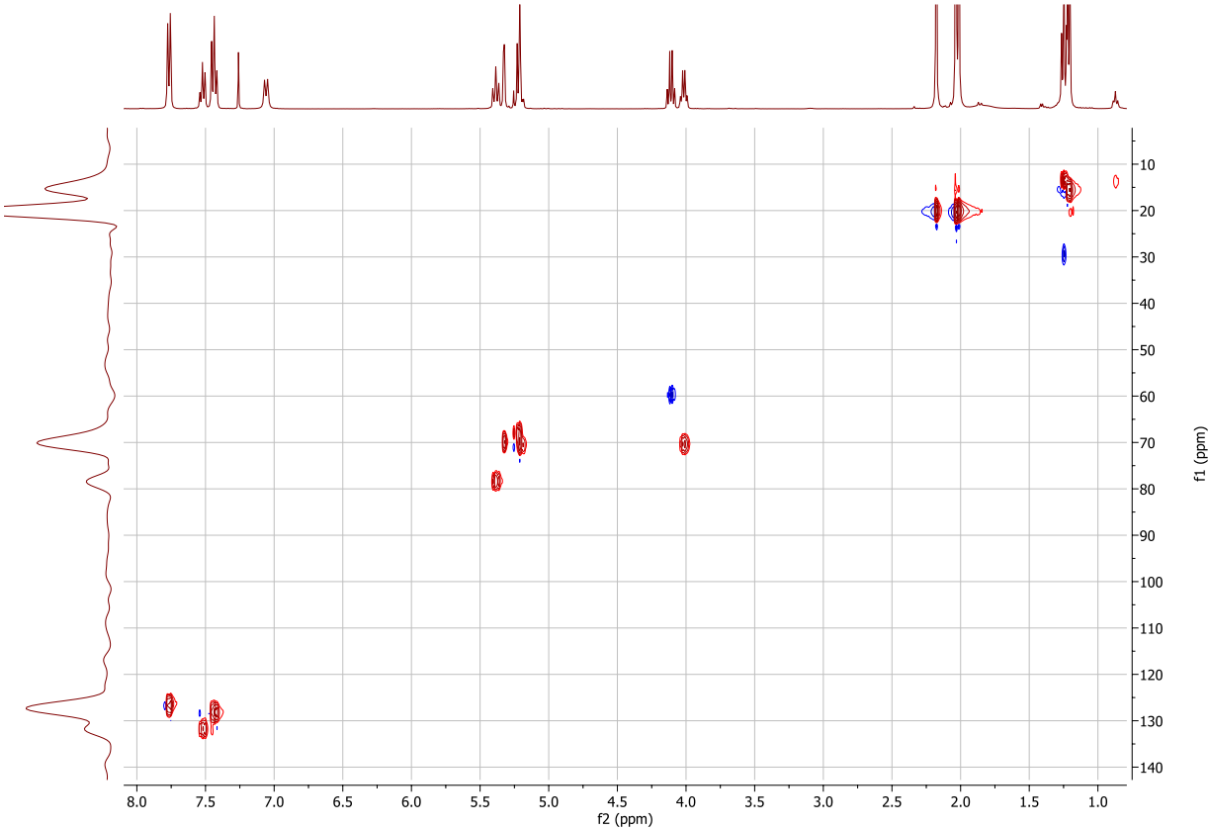
δ = 7.77 (d, J = 7.7 Hz, 2H, *CH Ar*), 7.52 (t, J' = 7.4 Hz, 1H, *CH Ar*), 7.44 (dd, J = 7.7 Hz, J' = 7.4 Hz, 2H, *CH Ar*), 7.06 (d, J_{NH-1} = 8.9 Hz, 1H, *NH*), 5.39 (t, J_{1-2} = J_{1-NH} = 8.9 Hz, 1H, *H-1*), 5.33 (dd, J_{4-3} = 2.9 Hz, J_{4-5} = 1.1 Hz, 1H, *H-4*), 5.22 (mult., 2H, *H-2 + H-3*), 4.02 (dq, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.4 Hz, 1H, *H-5*), 2.18, 2.03, 2.01 (s, 3H, *OAc*), 1.21 (d, J_{CH3-5} = 6.4 Hz, 3H, *CH_3*).



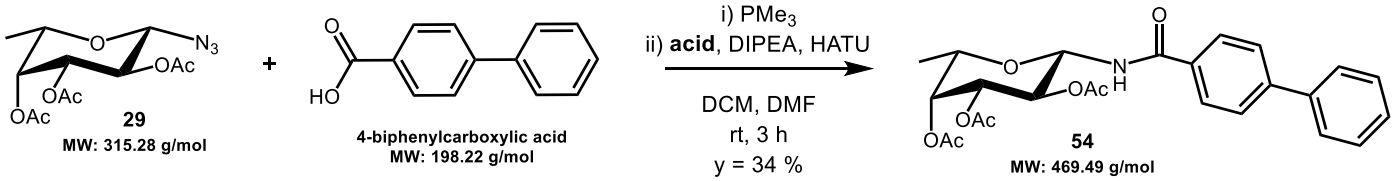
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 132.5, 128.9, 127.4 (*CH Ar*), 79.2 (C*1*), 71.4, 71.3 (*C3*), 71.1 (*C5*), 70.6 (*C4*), 68.7 (*C2*), 20.8 (*CH_3 OAc*), 16.3 (*C6*).

HSQC:



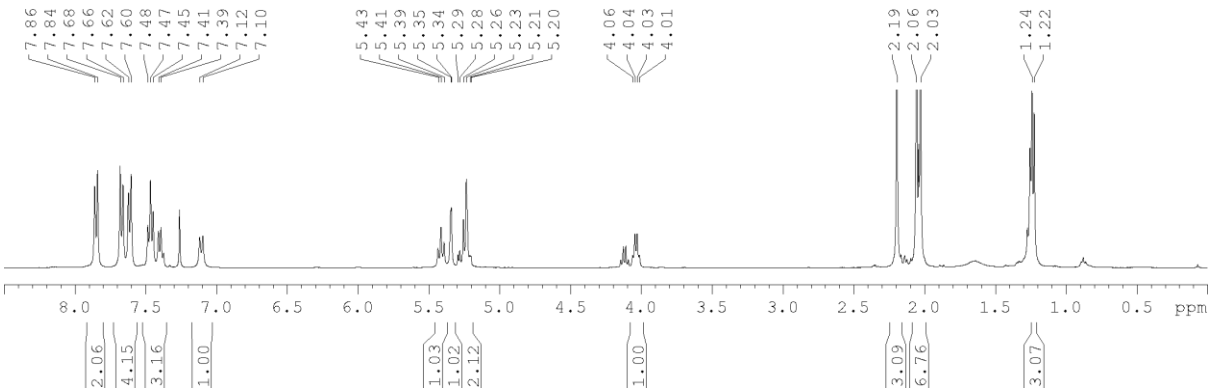
(54) Synthesis and characterization of **N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-[1,1'-biphenyl]-4-carboxamide (54)**:



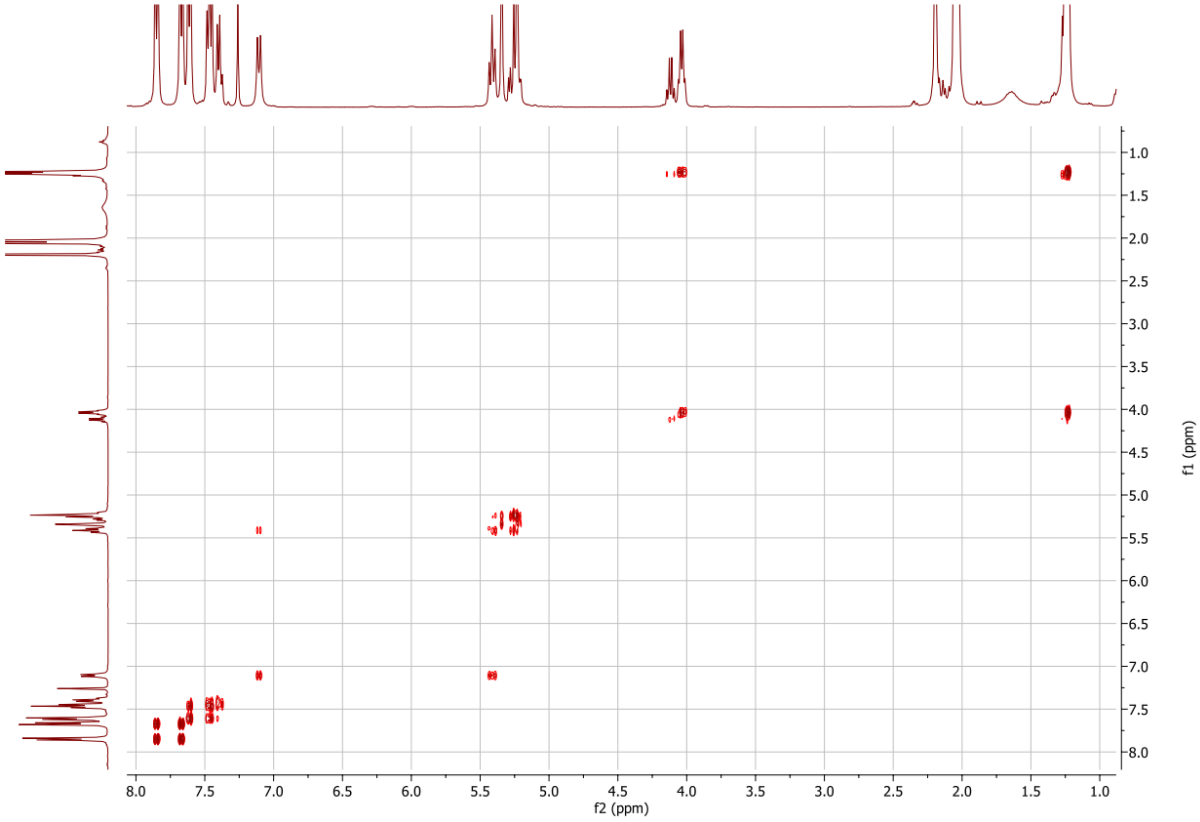
(2,3,4-tri-O-acetyl β-L-fucopyranosyl) azide 29 (0.127 mmol) was coupled to **4-biphenylcarboxylic acid** (0.190 mmol) following the protocol described, affording **54** (0.043 mmol, y = 34 %). TLC R_f (nHex/EtOAc: 6/4): 0.45.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.85 (d, J = 8.2 Hz, 2H, *CH Ar*), 7.64 (mult., 4H, *CH Ar*), 7.50 - 7.36 (mult., 3H, *CH Ar*), 7.11 (d, J_{NH-1} = 8.9 Hz, 1H, *NH*), 5.41 (t, J_{1-2} = J_{1-NH} = 8.9 Hz, 1H, *H-1*), 5.35 (dd, J_{4-3} = 2.9 Hz, J_{4-5} = 1.1 Hz, 1H, *H-4*), 5.35 (mult., 2H, *H-2 + H-3*), 4.04 (dq, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.4 Hz, 1H, *H-5*), 2.19, 2.06, 2.03 (s, 3H, *OAc*) 1.23 (d, J_{CH3-5} = 6.4 Hz, 3H, *CH_3*).

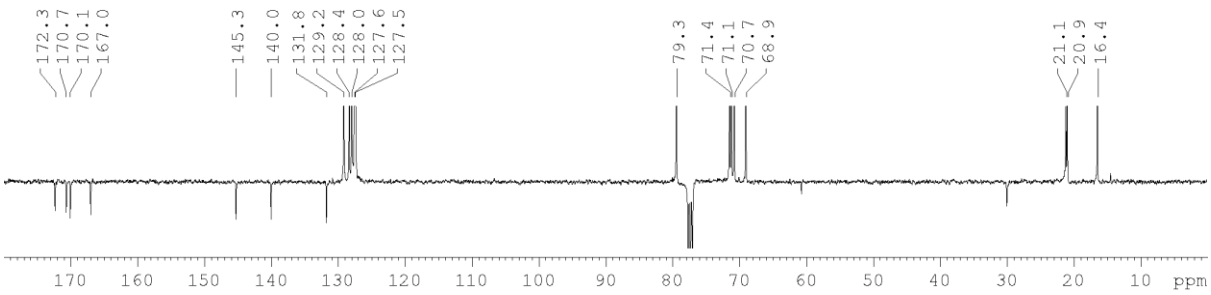


COSY:

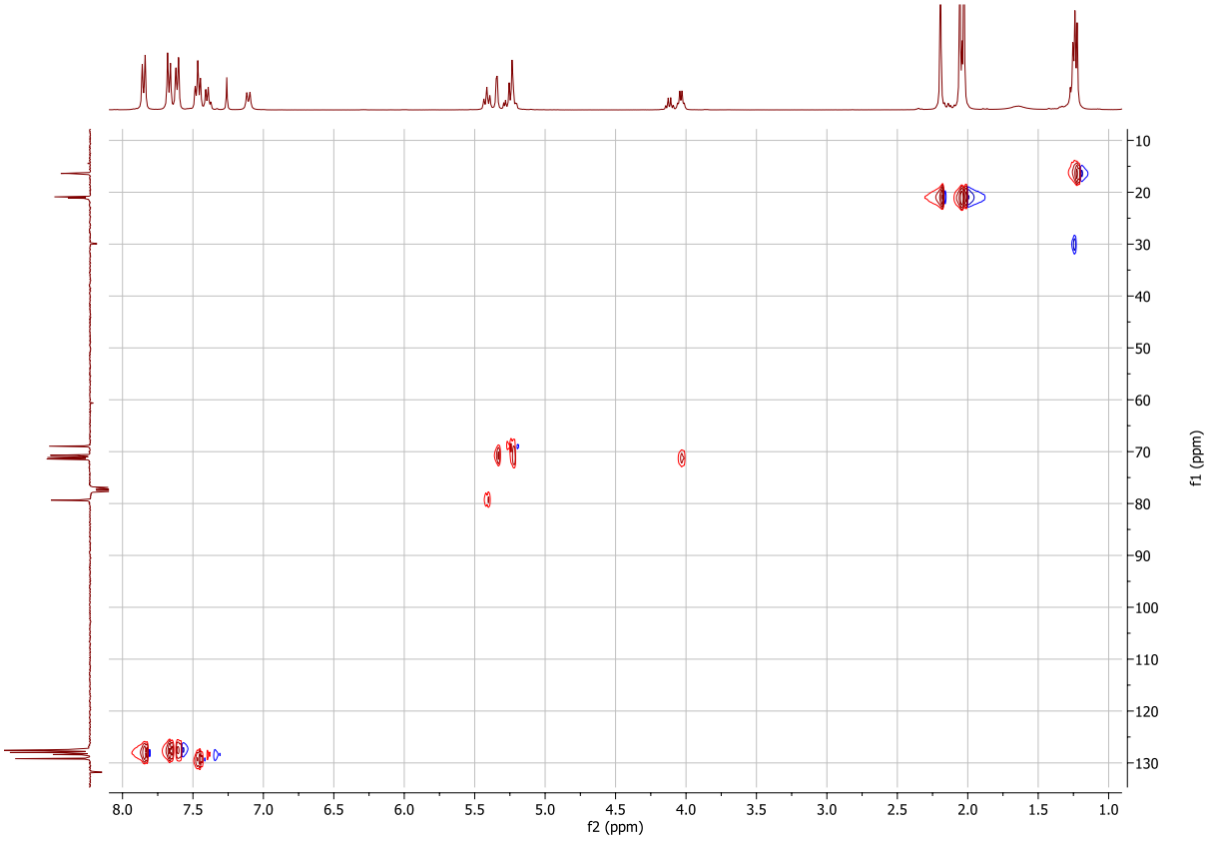


^{13}C NMR (400 MHz, $CDCl_3$):

δ = 172.3, 170.7, 170.1 (*C=O Ac*), 167.0 (*HN-C=O*), 145.3, 140.1, 131.8 (*C Ar*), 129.2, 128.4, 128.0, 127.6, 127.5 (*CH Ar*), 79.3 (C*1*), 71.4, 71.1 (*C3* + *C5*), 70.7 (*C4*), 68.9 (*C2*), 21.1, 20.9 (*CH_3 OAc*), 16.4 (*C6*).


172.3
170.7
170.1
167.0
145.3
140.0
131.8
129.2
128.4
128.0
127.6
127.5
79.3
71.4
71.1
70.7
68.9
21.1
20.9
16.4
170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 ppm

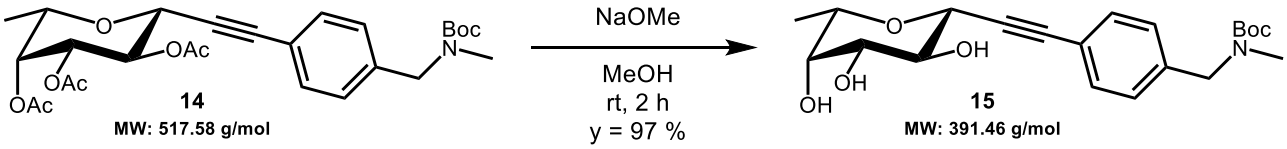
HSQC:


f1 (ppm)
f2 (ppm)

General procedure for Zemplén deacetylation, from Lowary and co-workers:[181]

To the **acetylated compound** (1.0 eq) in dry MeOH (concentration: 0.04 M) was added a 0.1 M solution of NaOMe in MeOH (2.5 eq). The reaction mixture was stirred at room temperature for 1 - 3 h until TLC showed completion, before being neutralized with Amberlite® IR_{120} H^+ resin, filtered, and concentrated. The crude was used for the following step directly or after purification by automatic chromatography (DCM/MeOH) or automatic reverse phase chromatography (H_2O/MeOH).

(15) Synthesis and characterization of ***tert*-butyl (4-(β-L-fucopyranosylethynyl)benzyl) (methyl) carbamate (15)**:



tert-butyl (4-(2,3,4-tri-O-acetyl β-L-fucopyranosylethynyl)benzyl) (methyl) carbamate 14 (0.097 mmol) was subjected to the aforementioned procedure to afford **15** (0.093 mmol, y = 97 %). TLC R_f (nHex/EtOAc: 1/1 and DCM/MeOH: 9/1): 0.03 and 0.39. MS (ESI) calculated for $C_{21}H_{29}NO_6$ $[M + Na]^+$ *m/z*: 414.19; found: 414.25. $[\alpha]_D^{17}$ = 0.5 (MeOH, c 1).

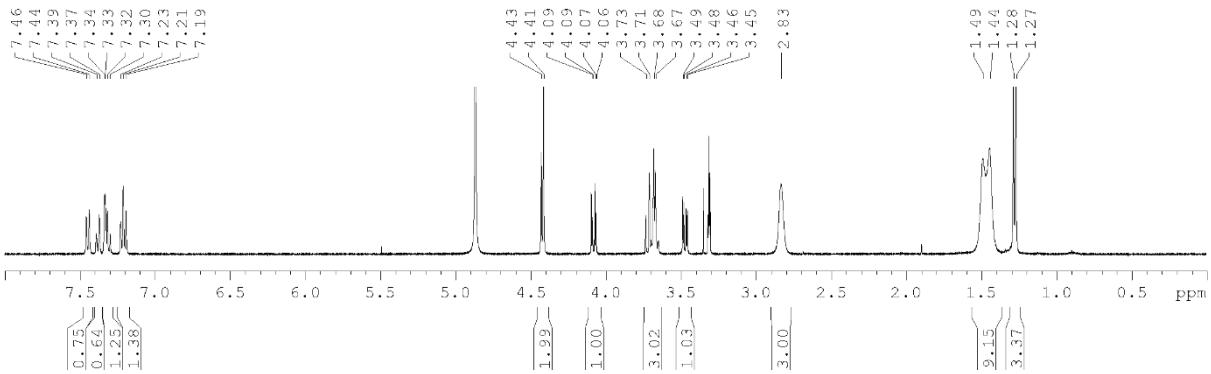
^{1}H NMR (400 MHz, MeOD): regioisomeric *para*/*meta* mixture (36/64)

para-isomer:

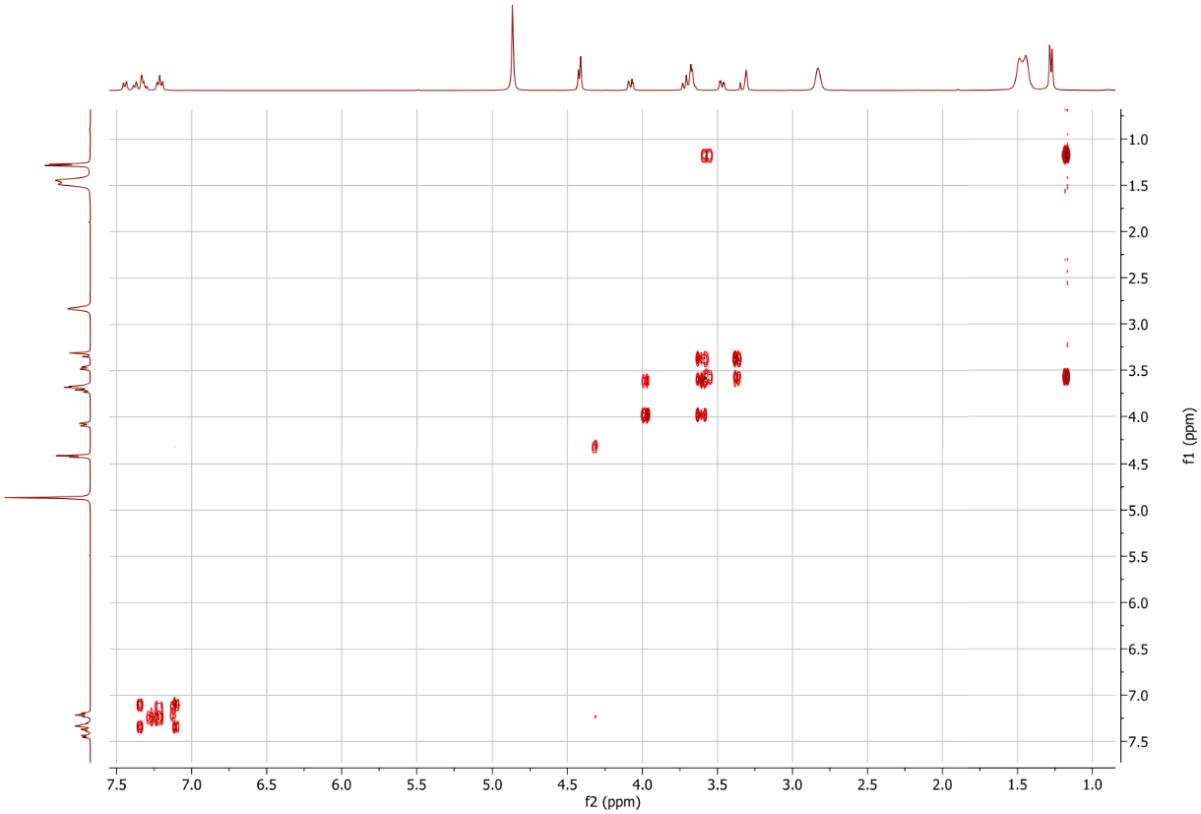
δ = 7.45 (d, J = 8.2 Hz, 2H, *CH Ar*), 7.21 (d, J = 8.0 Hz, 2H, *CH Ar*), 4.43 (bs, 2H, CH_2), 4.07 (d, J_{1-2} = 9.6 Hz, 1H, *H-1*), 3.69 (mult., 3H, *H-2* + *H-4* + *H-5*), 3.47 (dd, J_{3-2} = 9.5 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 2.83 (bd, 3H, *N-*CH_3), 1.49 (bs, 9H, *tBu*), 1.28 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).

meta-isomer:

δ = 7.38 (d, J = 7.6 Hz, 1H, *CH Ar*), 7.32 (mult., J = 7.7 Hz, 2H, *CH Ar*), 7.21 (m, 1H, *CH Ar*), 4.41 (bs, 2H, CH_2), 4.08 (d, J_{1-2} = 9.6 Hz, 1H, *H-1*), 3.69 (mult., J_{2-1} = J_{2-3} = 9.6 Hz, 3H, *H-2* + *H-4* + *H-5*), 3.47 (dd, J_{3-2} = 9.5 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 2.83 (bd, 3H, *N-*CH_3), 1.44 (bs, 9H, *tBu*), 1.28 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).

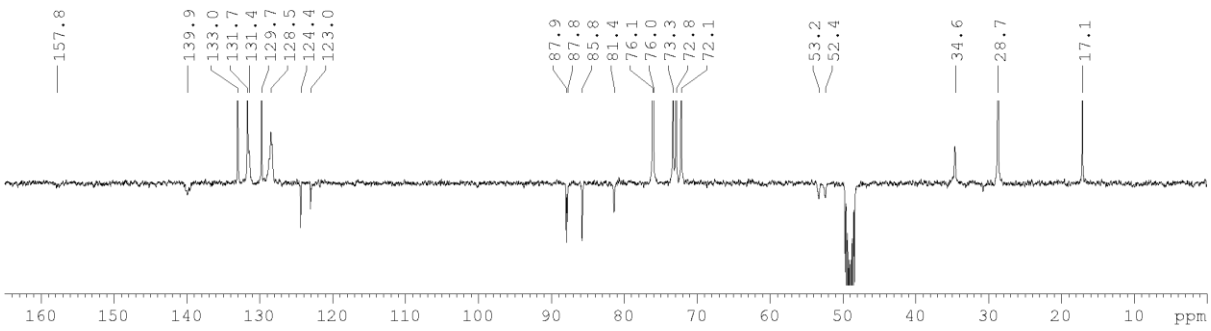


COSY:

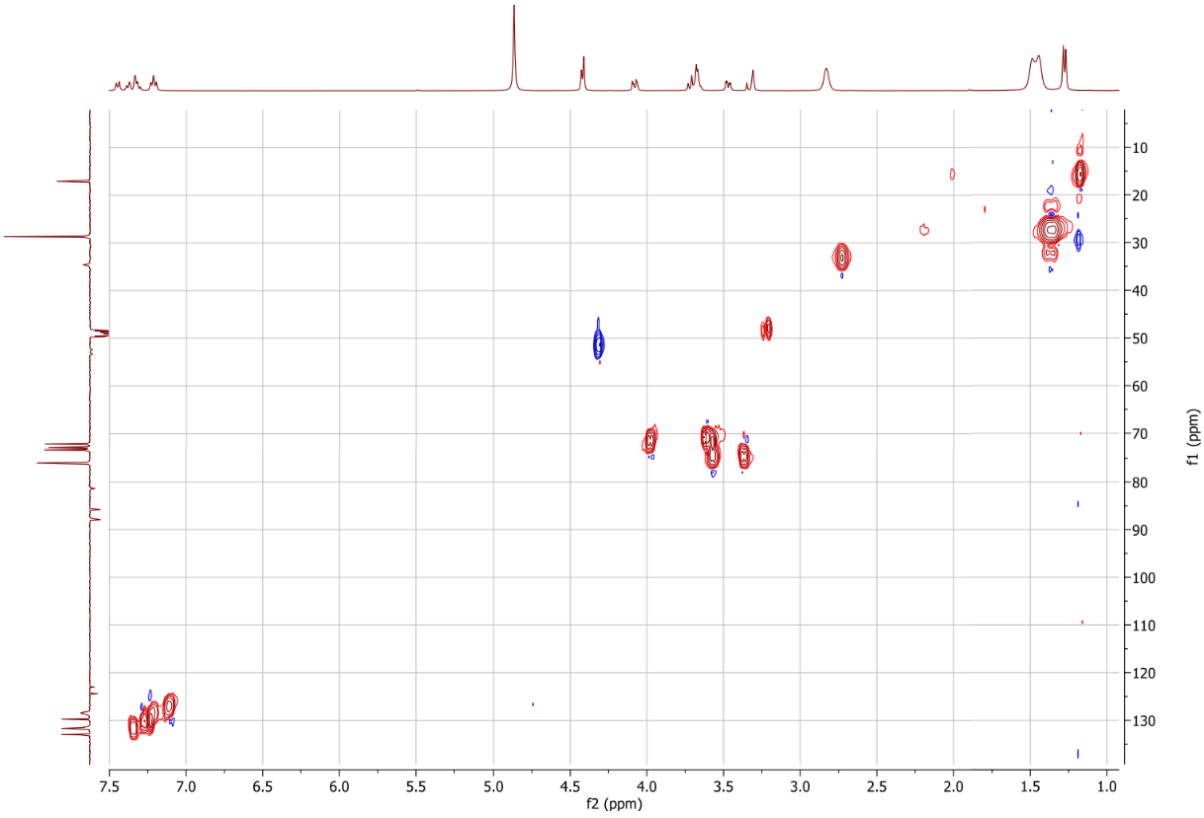


^{13}C NMR (400 MHz, MeOD):

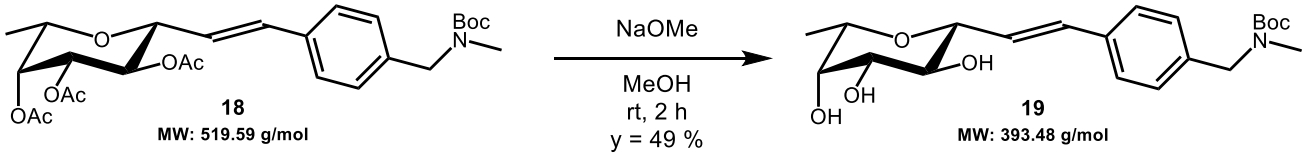
δ = 157.8 (*C=O tBu*), 139.9 (*CH_2-C Ar*), 133.0, 131.7, 129.7 (*meta CH Ar*), 131.4, 128.5 (*para CH Ar*), 124.4 (*meta ≡C-C Ar*), 123.0 (*para ≡C-C Ar*), 87.9 (*meta C1-C≡*), 87.8 (*para C1-C≡*), 85.8 (*≡C-Ar*), 81.4 (*C tBu*), 76.1, 76.0 (*C3 + C5*), 73.3 (*C2*), 72.8 (*C1*), 72.1 (*C4*), 53.2, 52.4 (*CH_2*), 34.6 (*CH_3*), 28.7 (*CH_3 tBu*), 17.1 (*C6*).



HSQC:



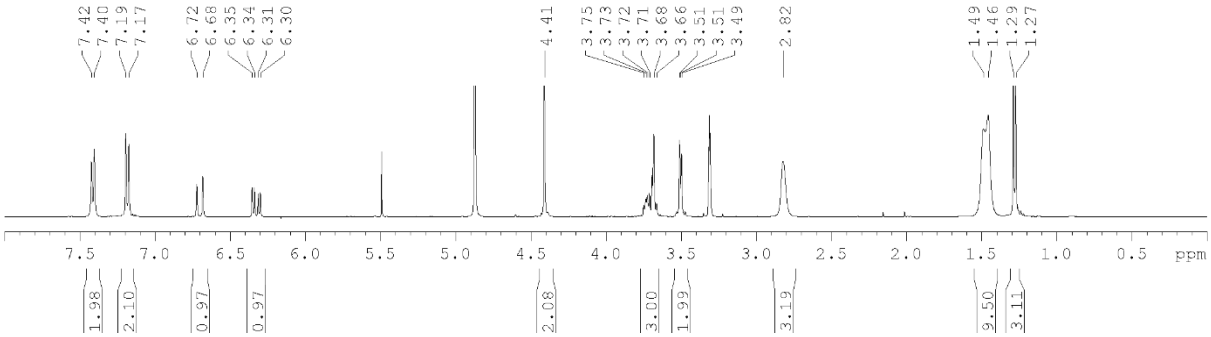
(19) Synthesis and characterization of ***tert*-butyl (*E*)-(4-(β-L-fucopyranosylvinyl)benzyl) (methyl) carbamate (19)**:



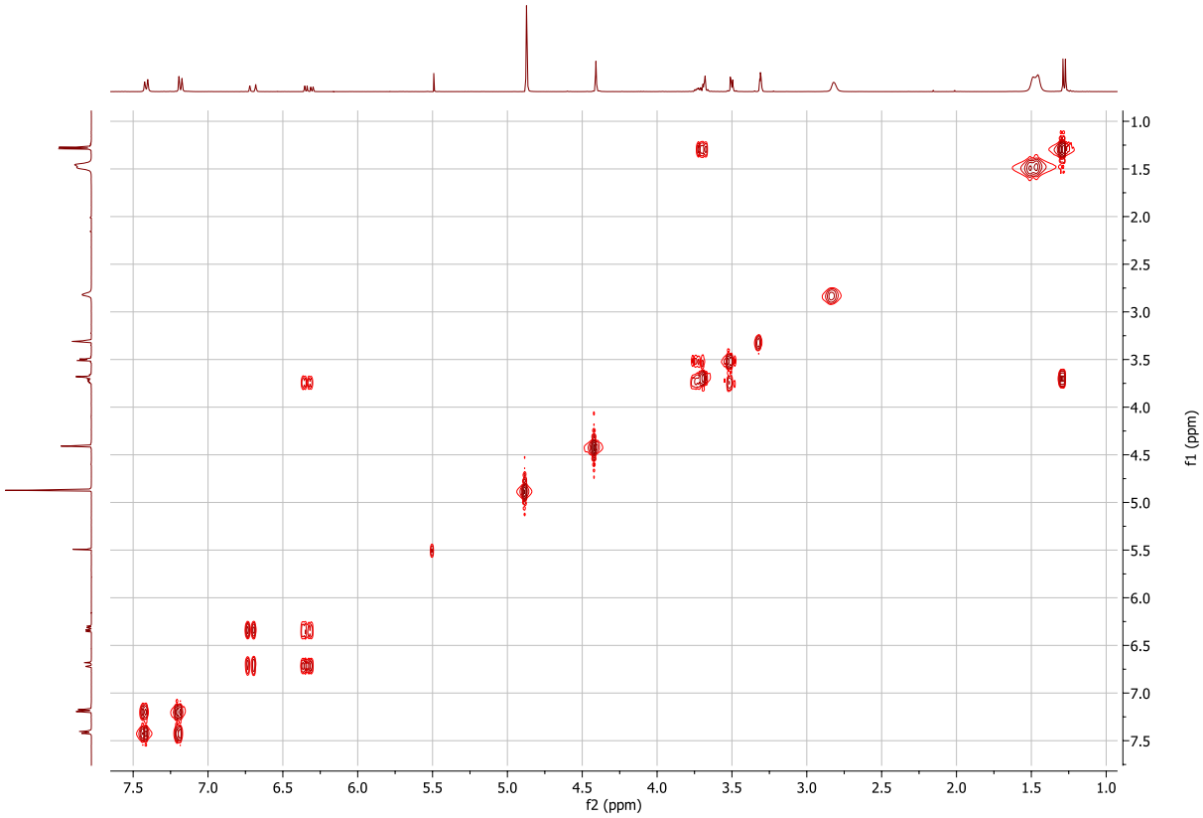
***tert*-butyl (*E*)-(4-(2,3,4-tri-O-acetyl β-L-fucopyranosylvinyl)benzyl) (methyl) carbamate 18** (0.067 mmol) was subjected to the aforementioned procedure to afford **19** (0.033 mmol, y = 49 %). TLC R_f (DCM/MeOH: 95/5): 0.18. MS (ESI) calculated for $C_{21}H_{31}NO_6$ $[M + Na]^+$ *m/z*: 416.20; found: 416.40.

^{1}H NMR (400 MHz, MeOD):

δ = 7.41 (d, J = 8.0 Hz, 2H, *Ar*), 7.18 (d, J = 8.1 Hz, 2H, *Ar*), 6.70 (d, J_{trans} = 16.0 Hz, 1H, =*CH*-*Ar*), 6.32 (dd, J_{trans} = 16.0 Hz, J_{CH-1} = 6.4 Hz, 1H, *C1*-*CH*=), 4.41 (bs, 2H, CH_2), 3.71 (mult., J_{5-4} = 1.3, Hz, J_{5-CH3} = 6.4 Hz, 3H, *H-1* + *H-3* + *H-5*), 3.50 (mult, J_{2-1}= 9.6 Hz, 2H, *H-2* + *H-4*), 2.82 (bd, 3H, *N-*CH_3), 1.47 (bs, 9H, *tBu*), 1.28 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).

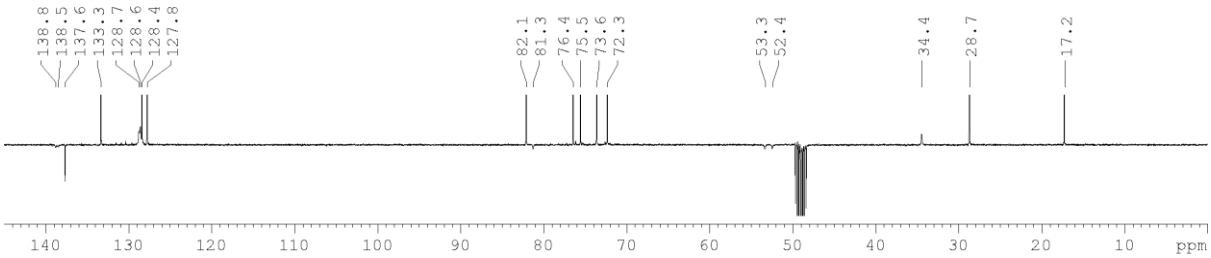


COSY:

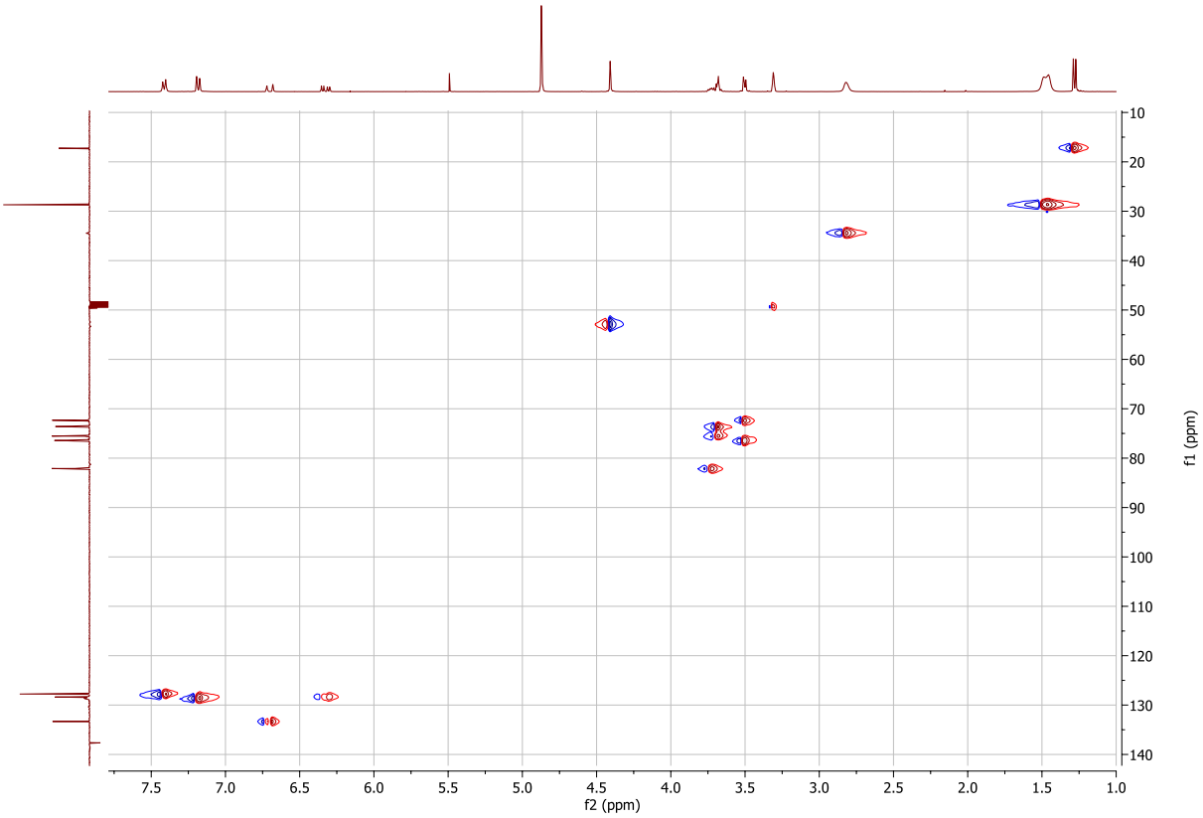


^{13}C NMR (400 MHz, MeOD):

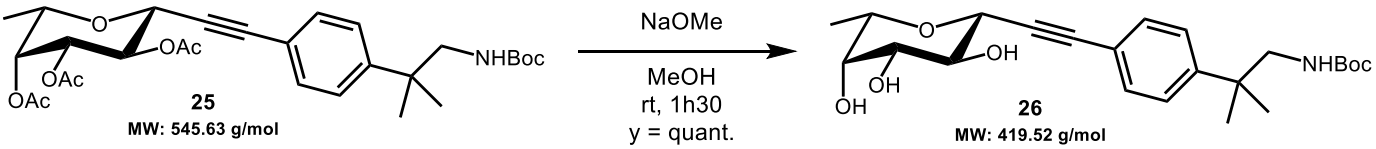
δ = 138.8 - 138.5, 137.6, (*C Ar*), 133.3 (*CH*-*Ar*), 128.7 - 128.6 (*CH Ar*), 128.4 (*C1*-*CH*=), 127.8 (*CH Ar*), 82.1 (*C5*), 81.3 (*C tBu*), 76.4 (*C2*), 75.5 (*C3*), 73.6 (*C1*), 72.3 (*C4*), 53.3, 52.4 (CH_2), 34.4 (CH_3), 28.7 (CH_3 *tBu*), 17.2 (*C6*).



HSQC:



(26) Synthesis and characterization of ***tert*-butyl (2-(4-(β-L-fucopyranosylethynyl)phenyl)-2-methylpropyl)carbamate (26)**:



***tert*-butyl (2-(4-(2,3,4-tri-O-acetyl β-L-fucopyranosylethynyl)phenyl)-2-methylpropyl) carbamate 25** (0.051 mmol) was subjected to the aforementioned procedure to afford **15** (quantitative yield). TLC R_f (nHex/EtOAc: 1/1 and DCM/MeOH: 9/1): 0.03 and 0.40.

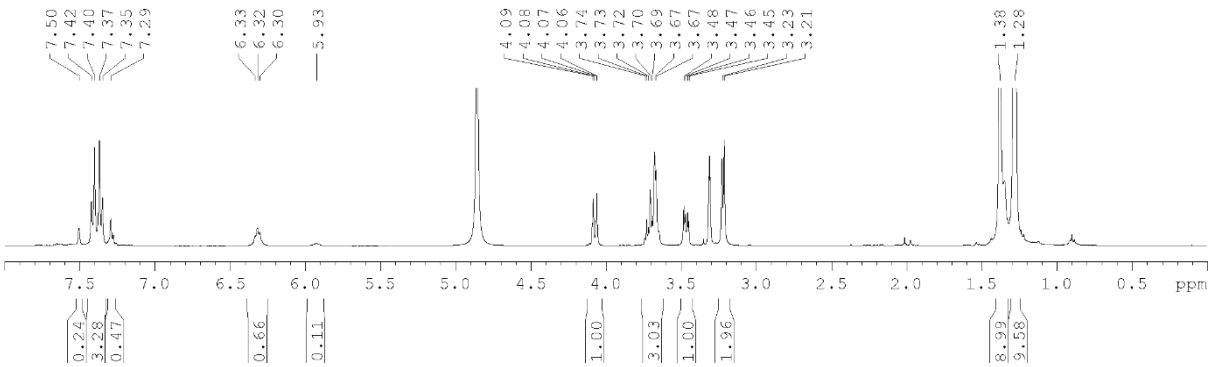
^{1}H NMR (400 MHz, MeOD): regioisomeric *para*/*meta* mixture (75/25)

para-isomer:

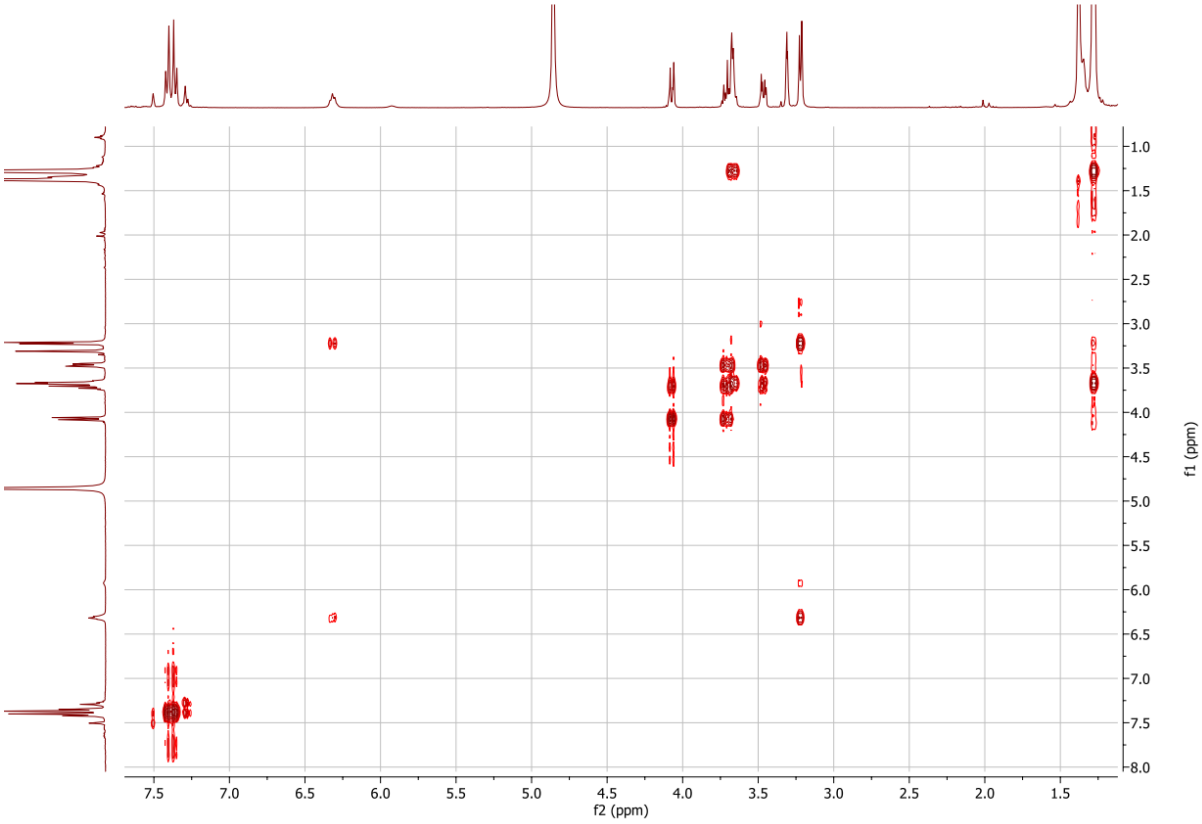
δ = 7.41, 7.36 (d, J = 8.4 Hz, 4H, *CH Ar*), 6.32 (bt, J = 6.3 Hz, 1H, *NH*), 4.07 (d, $J_{1\text{-}2}$ = 9.6 Hz, 1H, *H-1*), 3.70 (mult., $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.5 Hz, 3H, *H-2* + *H-4* + *H-5*), 3.46 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.3 Hz, 1H, *H-3*), 3.21 (bs, 2H, CH_2), 1.38 (bs, 9H, *tBu*), 1.28 (mult., 9H, 2xCH_3 + 3x*H-6*).

meta-isomer:

δ = 7.50 (bs, 1H, *CH Ar*), 7.38 (m, 1H, *CH Ar*), 7.29 (mult., 2H, *CH Ar*), 5.93 (bs, 1H, *NH*), 4.08 (d, $J_{1\text{-}2}$ = 9.6 Hz, 1H, *H-1*), 3.72 (mult., $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.5 Hz, 3H, *H-2* + *H-4* + *H-5*), 3.48 (dd, $J_{3\text{-}2}$ = 9.4 Hz, 1H, *H-3*), 3.21 (bs, 2H, CH_2), 1.38 (bs, 9H, *tBu*), 1.28 (mult., 9H, 2xCH_3 + 3x*H-6*).

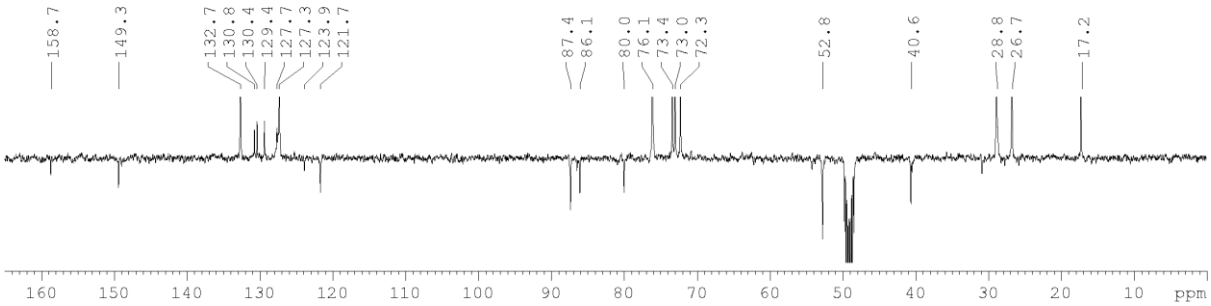


COSY:

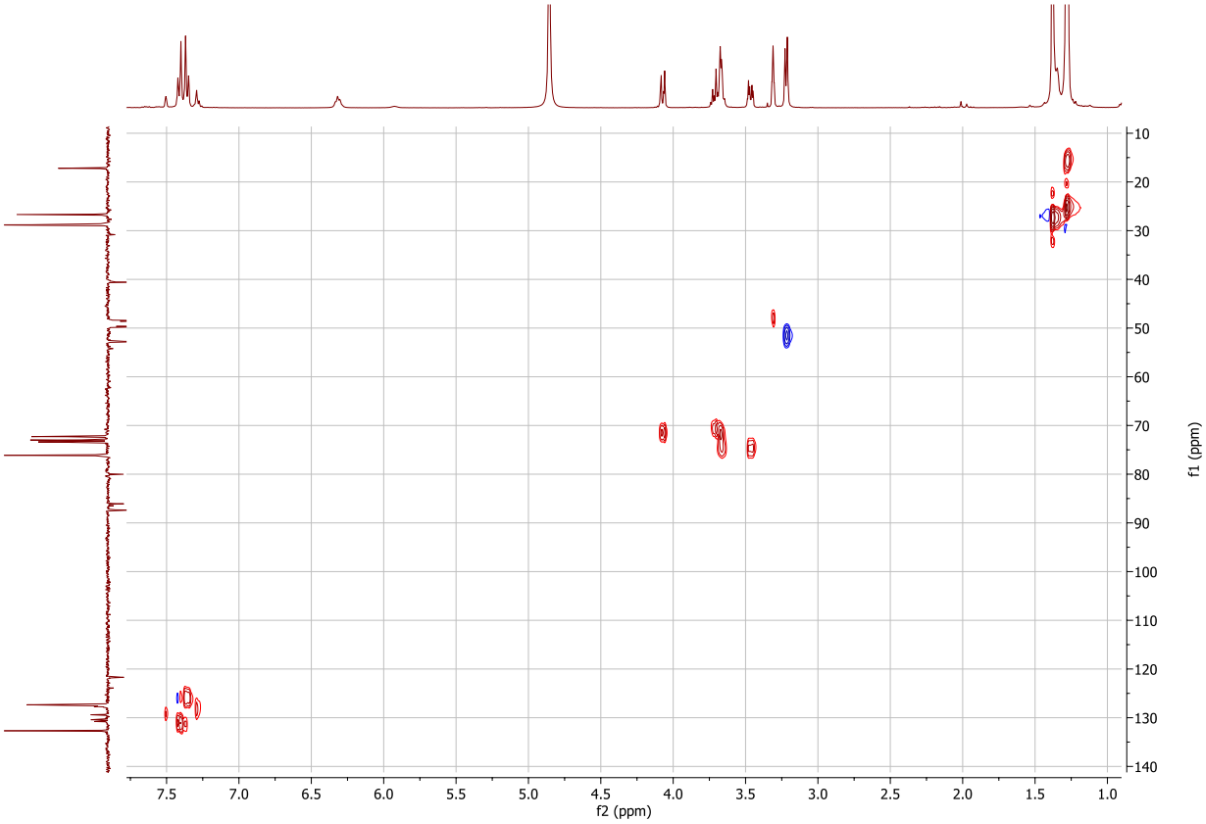


^{13}C NMR (400 MHz, MeOD):

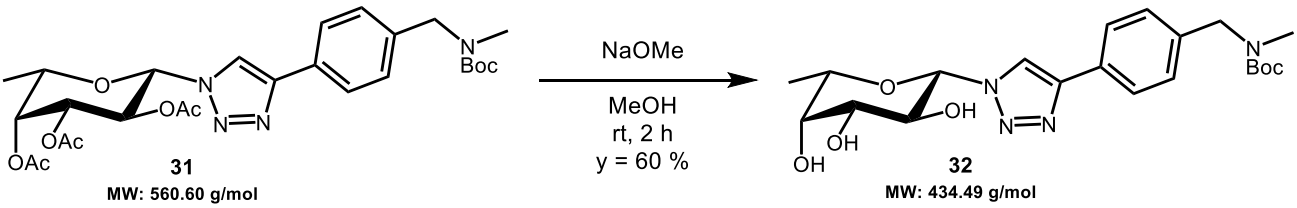
δ = 158.7 (*C=O tBu*), 149.3 (*C(CH₃)₂-C Ar*), 132.7 (*para CH Ar*), 130.8, 130.4, 129.4, 127.7 (*meta CH Ar*), 127.3 (*para CH Ar*), 123.9 (*meta ≡C-C Ar*), 121.7 (*para ≡C-C Ar*), 87.4 (*C1-C≡*), 86.1 (*≡C-Ar*), 80.0 (*C tBu*), 76.1 (*C3 + C5*), 73.4 (*C4*), 73.0 (*C1*), 72.3 (*C2*), 52.8 (CH_2), 40.6 ($C(CH_3)_2$), 28.8 (CH_3 *tBu*), 26.7 (2xCH_3), 17.2 (*C6*).



HSQC:



(32) Synthesis and characterization of ***tert*-butyl (4-(1-(β-L-fucopyranosyl)-1H-1,2,3-triazol-4-**

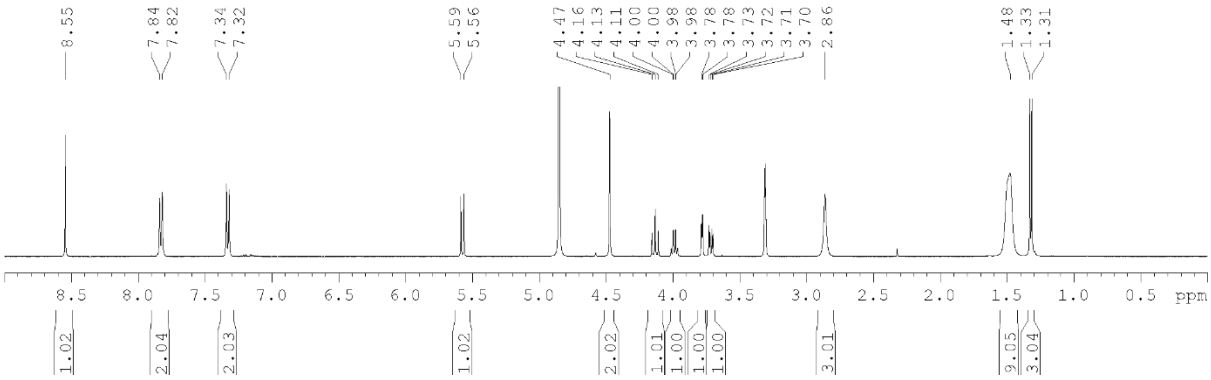


yl)benzyl) (methyl) carbamate (32):

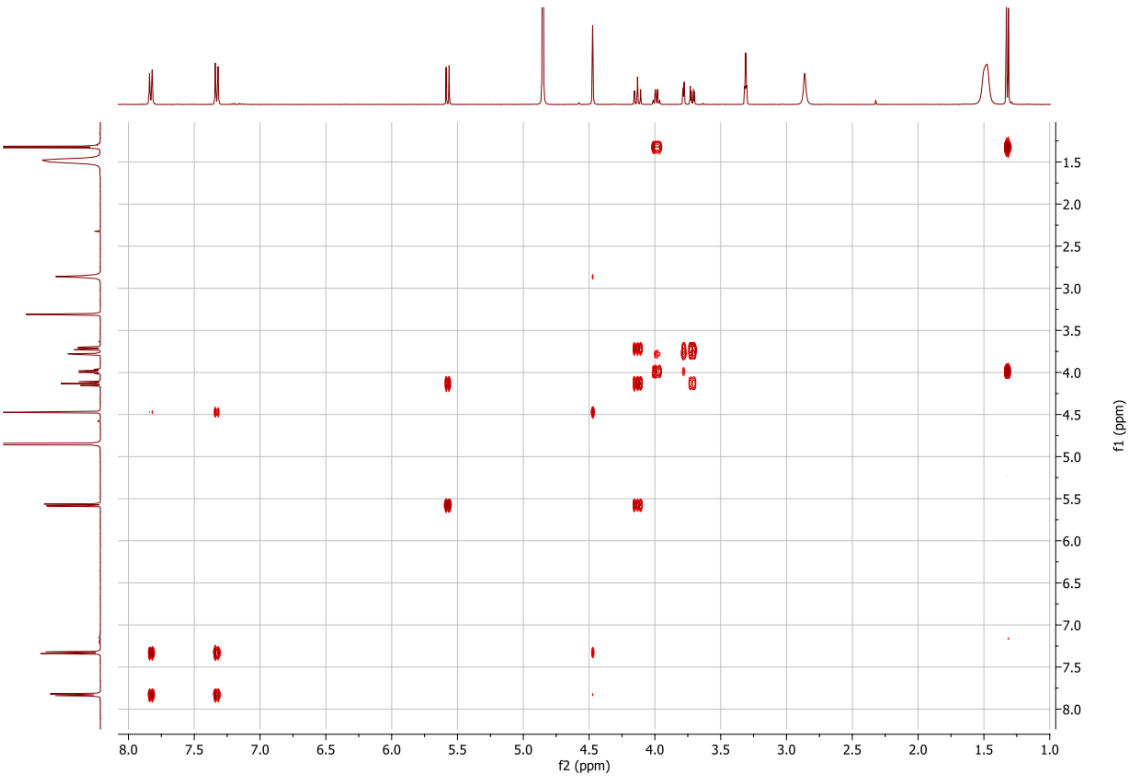
***tert*-butyl (4-(1-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)benzyl) (methyl) carbamate 31** (0.143 mmol) was subjected to the aforementioned procedure to afford **32** (0.086 mmol, y = 60 %). TLC R_f (DCM/MeOH: 85/15): 0.48. MS (ESI) calculated for $C_{21}H_{30}N_4O_6$ $[M + Na]^+$ *m/z*: 457.21; found: 457.22. $[\alpha]_D^{17}$ = 6.44 (MeOH, c 1).

^{1}H NMR (400 MHz, MeOD):

δ = 8.55 (s, 1H, *CH hAr*), 7.83 (d, J = 8.1 Hz, 2H, *CH Ar*), 7.33 (d, J = 8.1 Hz, 2H, *CH Ar*), 5.57 (d, $J_{1\text{-}2}$ = 9.2 Hz, 1H, *H-1*), 4.47 (bs, 2H, *CH_2*), 4.13 (dd, $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.3 Hz, 1H, *H-2*), 3.99 (qd, $J_{5\text{-}4}$ = 1.1 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.78 (dd, $J_{4\text{-}3}$ = 3.4 Hz, $J_{4\text{-}5}$ = 1.1 Hz, 1H, *H-4*), 3.72 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.3 Hz, 1H, *H-3*), 2.86 (bd, 3H, *N-CH_3*), 1.48 (bs, 9H, *tBu*), 1.32 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*).

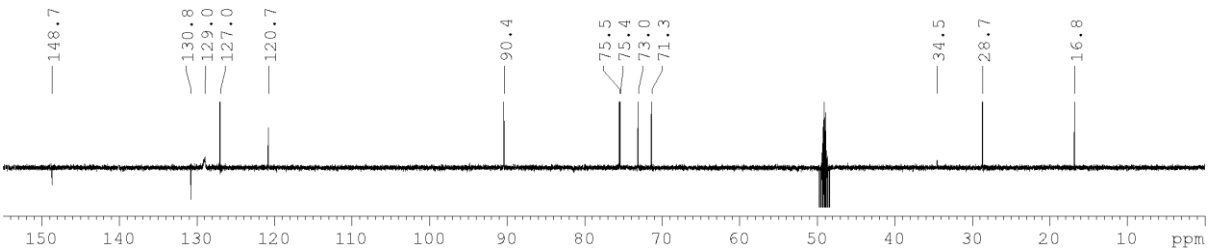


COSY:

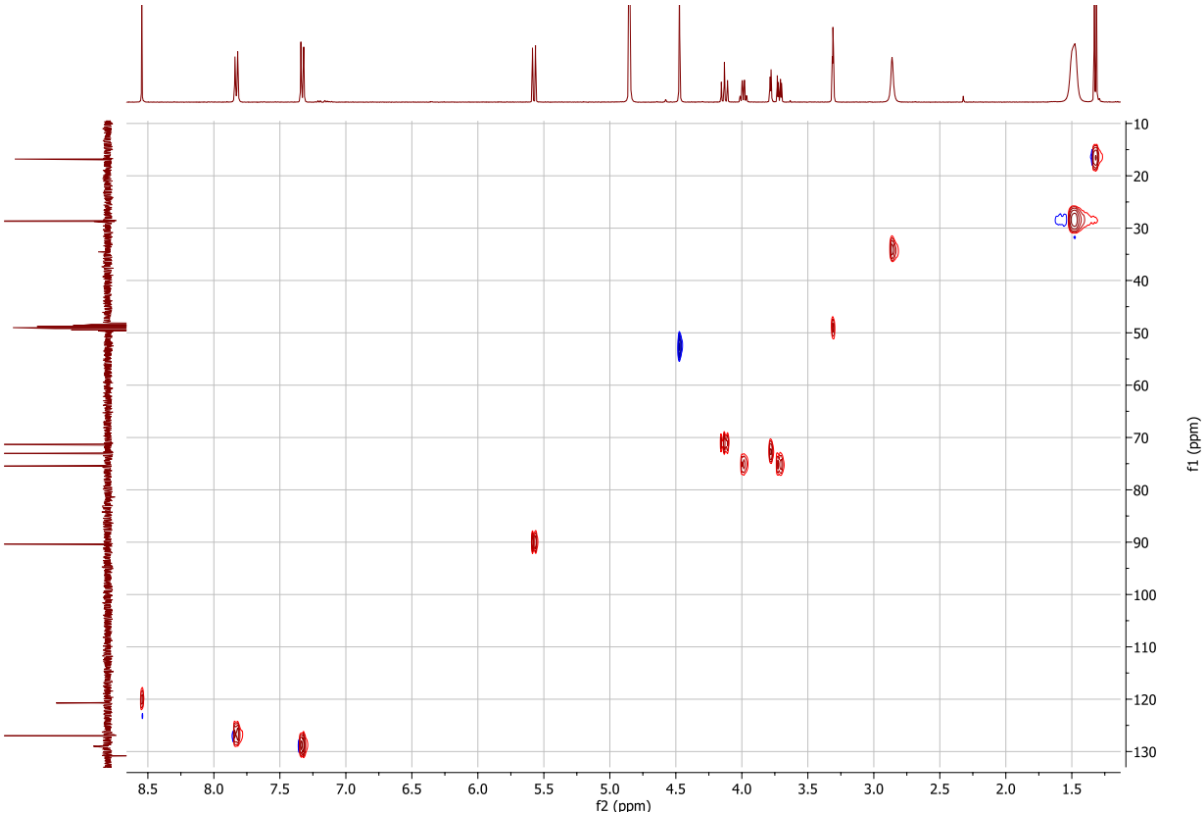


^{13}C NMR (400 MHz, MeOD):

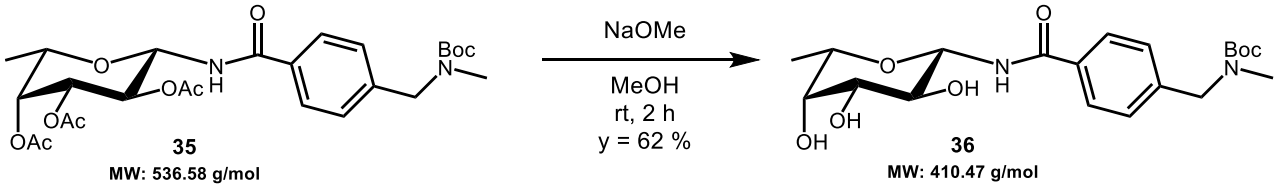
δ = 148.7 (*C=O tBu*), 130.8 (*C hAr*), 129.0, 127.0 (*CH Ar*), 120.7 (*CH hAr*), 90.4 (*C1*), 75.5, 75.4 (*C3, C5*), 73.0 (*C4*), 71.3 (*C2*), 52.7 (CH_2), 34.5 (CH_3), 28.7 (CH_3 *tBu*), 16.8 (*C6*).



HSQC:



(36) Synthesis and characterization of ***tert*-butyl (4-((β-L-fucopyranosyl)carbamoyl)benzyl) (methyl) carbamate (36)**:

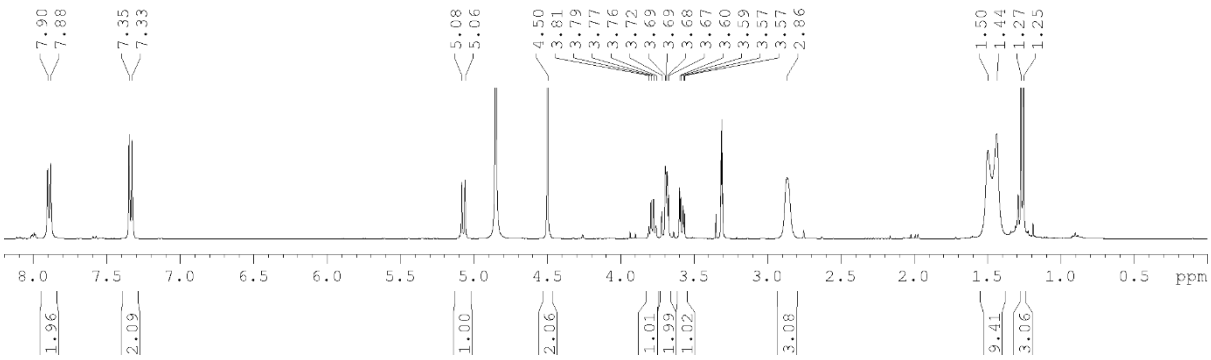


***tert*-butyl (4-((2,3,4-tri-O-acetyl β-L-fucopyranosyl)carbamoyl)benzyl) (methyl) carbamate 35** (0.086 mmol) was subjected to the aforementioned procedure to afford **36** (0.054 mmol,

y = 62 %). TLC R_f (DCM/MeOH: 95/5): 0.14. MS (ESI) calculated for $C_{20}H_{30}N_2O_7$ $[M + Na]^+$ *m/z*: 433.19; found: 433.32. $[\alpha]_D^{17.1}$ = -7.33 (MeOH, c 1).

^{1}H NMR (400 MHz, MeOD):

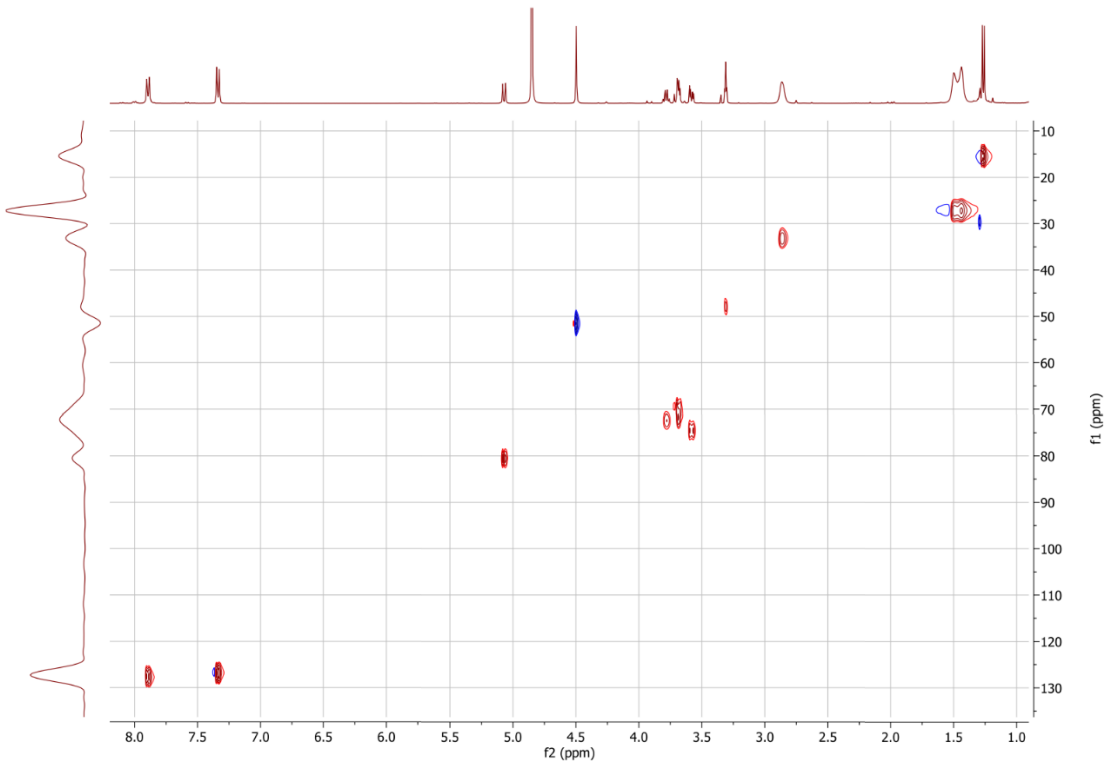
δ = 7.89 (d, J = 8.3 Hz, 2H, *CH Ar*), 7.34 (d, J = 8.6 Hz, 2H, *CH Ar*), 5.07 (d, J_{1-2} = 9.0 Hz, 1H, *H-1*), 4.50 (bs, 2H, *CH_2*), 3.78 (qd, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 3.69 (mult., J_{2-1} = J_{2-3} = 9.3 Hz, J_{4-3} = 3.4 Hz, J_{4-5} = 1.1 Hz, 2H, *H-2* + *H-4*), 3.58 (dd, J_{3-2} = 9.6 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 2.86 (bd, 3H, *N-CH_3*), 1.50, 1.48 (bs, 9H, *tBu*), 1.26 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*).



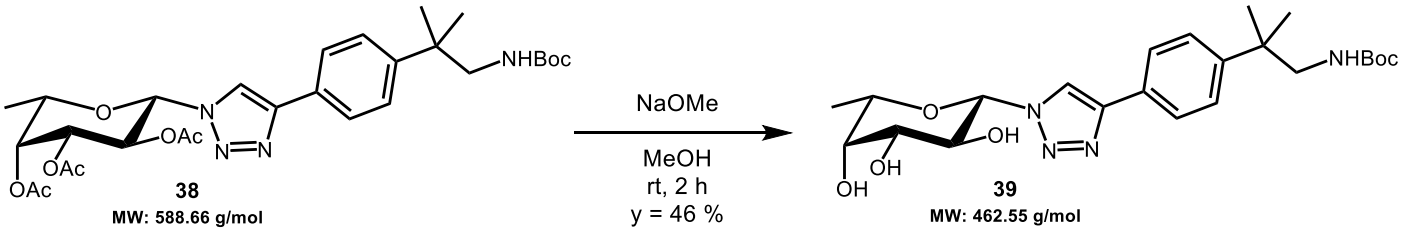
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 128.7, 127.9 (*CH Ar*), 81.7 (*C1*), 75.7 (*C3*), 73.5 (*C5*), 72.9 (*C4*), 70.6 (*C2*), 52.6 (*CH_2*), 34.3 (*CH_3*), 28.3 (*CH_3 tBu*), 16.5 (*C6*).

HSQC:



(39) Synthesis and characterization of ***tert*-butyl (2-(4-(1-(β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)phenyl)-2-methylpropyl) carbamate (39)**:



***tert*-butyl (2-(4-(1-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)phenyl)-2-methylpropyl) carbamate 38** (0.157 mmol) was subjected to the aforementioned procedure to afford **39** (0.063 mmol, y = 46 %). TLC R_f (DCM/MeOH: 95/5): 0.20. MS (ESI) calculated for $C_{23}H_{34}N_4O_6$ $[M + Na]^+$ *m/z*: 485.24; found: 485.58.

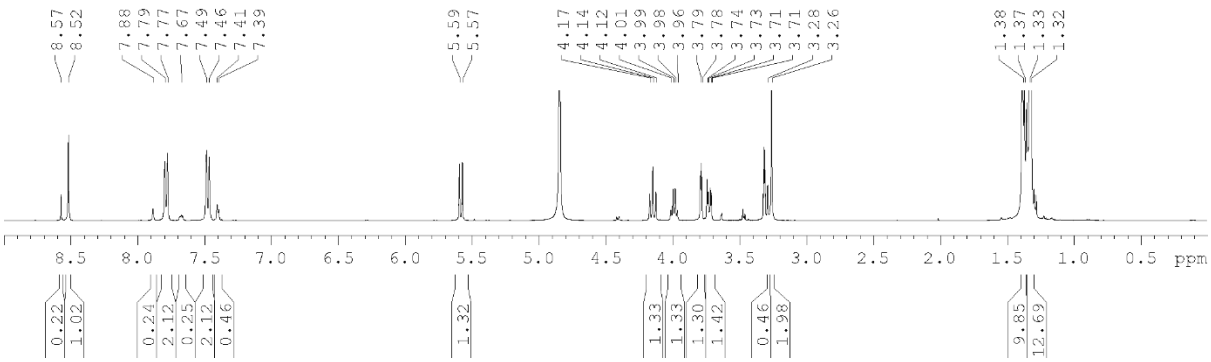
^{1}H NMR (400 MHz, MeOD): regioisomeric *para*/*meta* mixture (81/19)

para-isomer:

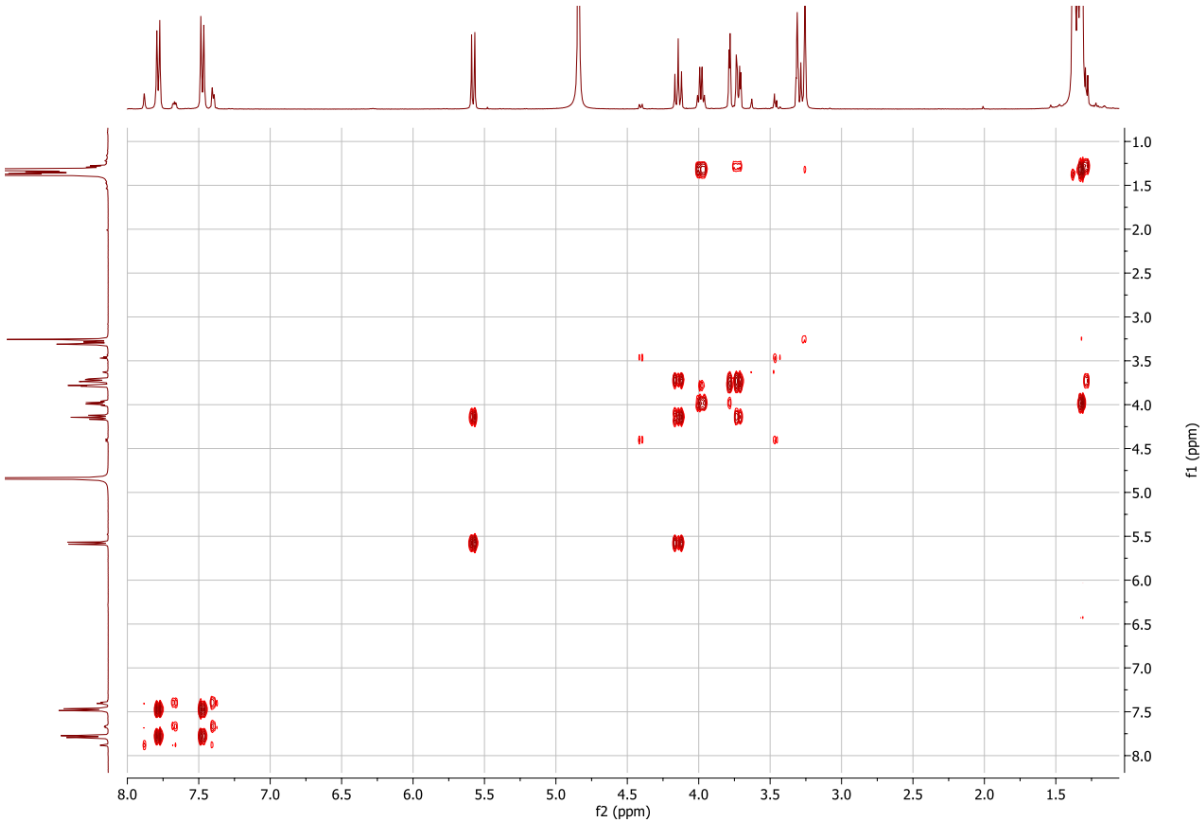
δ = 8.52 (s, 1H, *CH hAr*), 7.78 (d, J = 8.4 Hz, 2H, *CH Ar*), 7.47 (d, J = 8.5 Hz, 2H, *CH Ar*), 5.58 (d, $J_{1\text{-}2}$ = 9.2 Hz, 1H, *H-1*), 4.14 (dd, $J_{2\text{-}1}$ = 9.3 Hz, $J_{2\text{-}3}$ = 9.4 Hz, 1H, *H-2*), 3.99 (dq, $J_{5\text{-}4}$ = 1.2 Hz, $J_{5\text{-}6}$ = 6.5 Hz, 1H, *H-5*), 3.78 (d, $J_{4\text{-}3}$ = 3.3 Hz, 1H, *H-4*), 3.72 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.4 Hz, 1H, *H-3*), 3.26 (bs, 2H, CH_2), 1.37 (bs, 9H, *tBu*), 1.32 (mult., 9H, $J_{6\text{-}5}$ = 6.4 Hz, 2xCH_3 + 3x*H-6*).

meta-isomer:

δ = 8.57 (s, 1H, *CH hAr*), 7.88 (bs, 1H, *CH Ar*), 7.67 (m, 1H, *CH Ar*), 7.40 (mult., 2H, *CH Ar*), 5.58 (d, $J_{1\text{-}2}$ = 9.2 Hz, 1H, *H-1*), 4.14 (dd, 1H, *H-2*), 3.99 (dq, 1H, *H-5*), 3.78 (d, 1H, *H-4*), 3.73 (m, 1H, *H-3*), 3.28 (bs, 2H, CH_2), 1.38 (bs, 9H, *tBu*), 1.32 (mult., 9H, 2xCH_3 + 3x*H-6*).



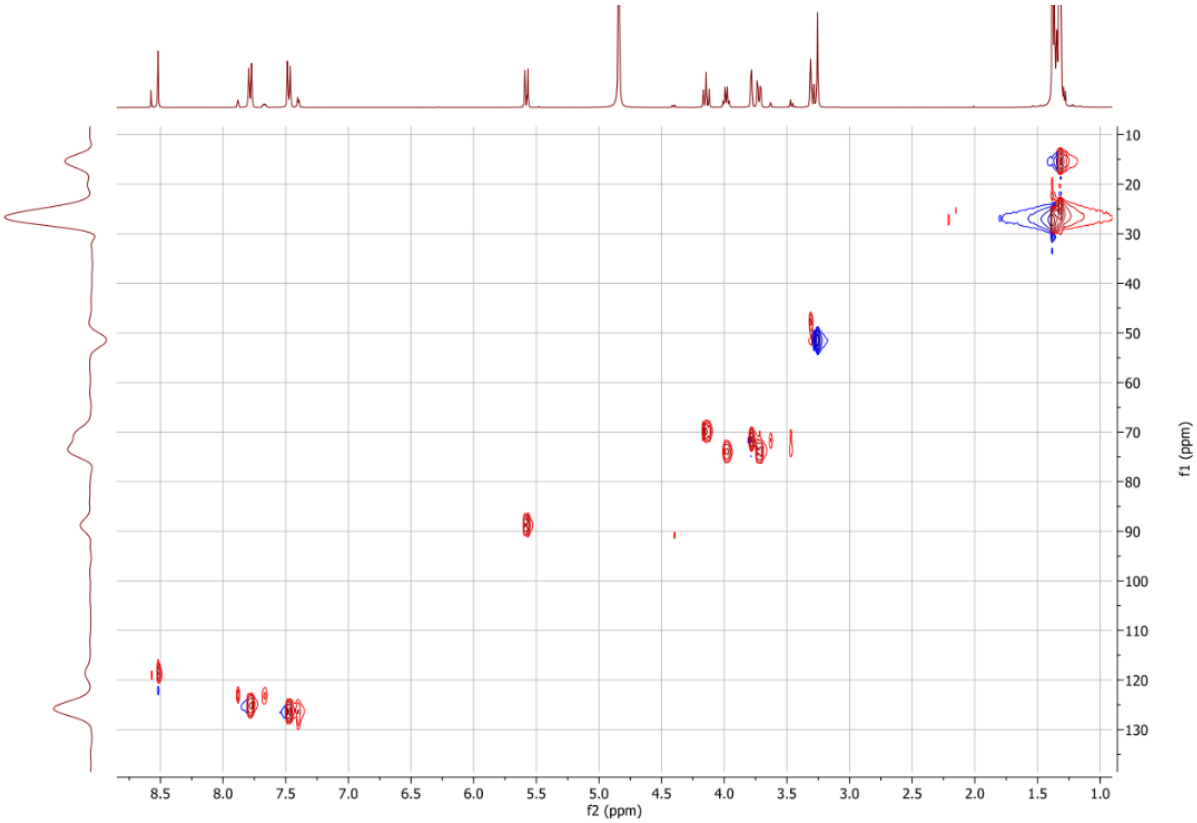
COSY:



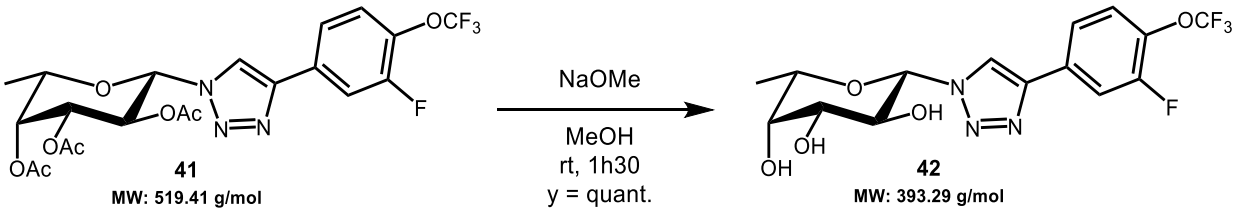
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 129.6, 127.1, 124.3, 124.0 (*meta CH Ar*), 127.4, 126.2 (*para CH Ar*), 120.2 (*meta CH hAr*), 120.0 (*para CH hAr*), 89.9 (*C1*), 75.1 (*C3*), 74.9 (*C5*), 72.6 (*C4*), 70.9 (*C2*), 52.5 (CH_2), 28.3 (CH_3 *tBu*), 26.3 (2xCH_3), 16.4 (*C6*).

HSQC:



(42) Synthesis and characterization of **4-(3-fluoro-4-(trifluoromethoxy)phenyl)-1-(β-L-fucopyranosyl)-1H-1,2,3-triazole (42)**:

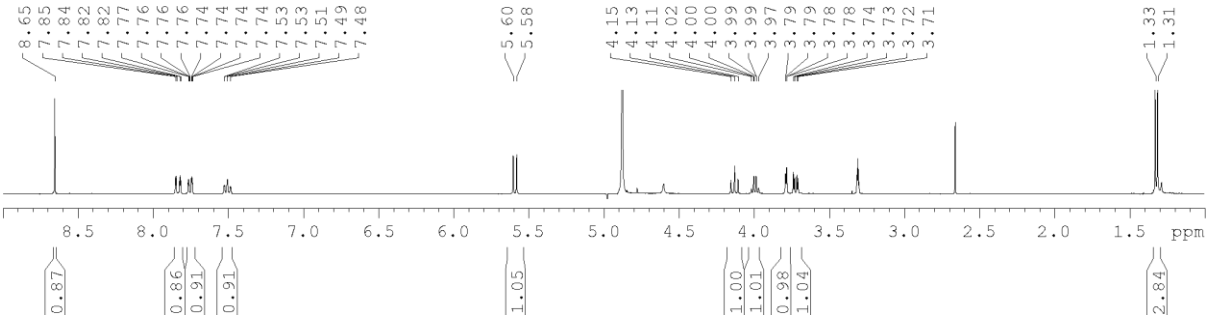


4-(3-fluoro-4-(trifluoromethoxy)phenyl)-1-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-1,2,3-triazole 41 (0.079 mmol) was subjected to the aforementioned procedure to afford **42** (quantitative yield). TLC R_f (EtOAc): 0.20. $[\alpha]_D^{23.7}$ = 6.7 (MeOH, c 0.3). MS (ESI) calculated for $C_{15}H_{15}F_4N_3O_5$ $[M + Na]^+$ *m/z*: 416.08; found: 416.01. HRMS (ESI^+-TOF) *m/z*: calculated for $C_{15}H_{15}F_4N_3O_5$ $[M + H]^+$: 394.1021, found: 394.1020.

^{1}H NMR (400 MHz, MeOD):

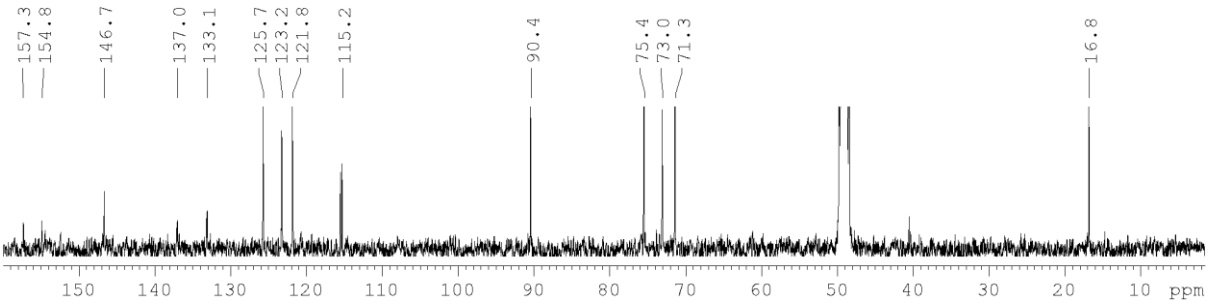
δ = 8.65 (s, 1H, *CH hAr*), 7.83 (dd, J = 11.3 Hz, J_{meta} = 2.1 Hz, 1H, *CH Ar*), 7.75 (ddd, J_{ortho} = 8.5 Hz, J_{meta} = 2.1 Hz, 1H, *CH Ar*), 7.51 (m, J_{ortho} = 7.9 Hz, J_{para} = 1.3 Hz, 1H, *CH Ar*), 5.59 (d, J_{1-2} = 9.2 Hz, 1H, *H-1*), 4.13 (dd, J_{2-1} = J_{2-3} = 9.3 Hz, 1H, *H-2*), 3.99 (dq, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.5 Hz, 1H, *H-*

5), 3.79 (d, J_{4-3} = 3.3 Hz, J_{4-5} = 1.1 Hz, 1H, *H-4*), 3.72 (dd, J_{3-2} = 9.5 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 1.32 (m, 3H, J_{CH3-5} = 6.5 Hz, *CH_3*).

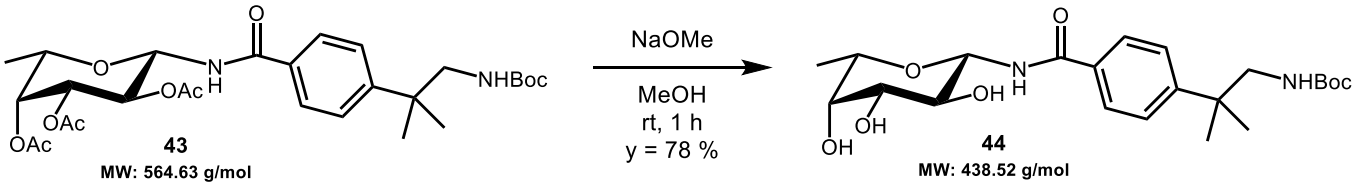


^{13}C NMR (400 MHz, MeOD):

δ = 157.4, 154.8, 146.7 (*CF*, *C-O*, *C hAr*), 137.0, 133.1 (*C Ar*, *CF_3*), 125.7 (*CH hAr*), 123.2, 121.8, 115.2 (*CH Ar*), 90.4 (*C1*), 75.5, 75.4 (*C3, C5*), 73.0 (*C4*), 71.3 (*C2*), 16.8 (*C6*).



(44) Synthesis and characterization of ***tert*-butyl (2-(4-((β-L-fucopyranosyl)carbamoyl)phenyl)-2-methylpropyl) carbamate (44)**:



***tert*-butyl (2-(4-((2,3,4-tri-O-acetyl β-L-fucopyranosyl)carbamoyl)phenyl)-2-methylpropyl) carbamate 43** (0.076 mmol) was subjected to the aforementioned procedure to afford **44** (0.059 mmol, y = 78 %). TLC R_f (DCM/MeOH: 95/5): 0.13. MS (ESI) calculated for $C_{22}H_{34}N_2O_7$ $[M + Na]^+$ *m/z*: 461.23; found: 461.28.

^{1}H NMR (400 MHz, MeOD): regioisomeric *para*/*meta* mixture (80/20)

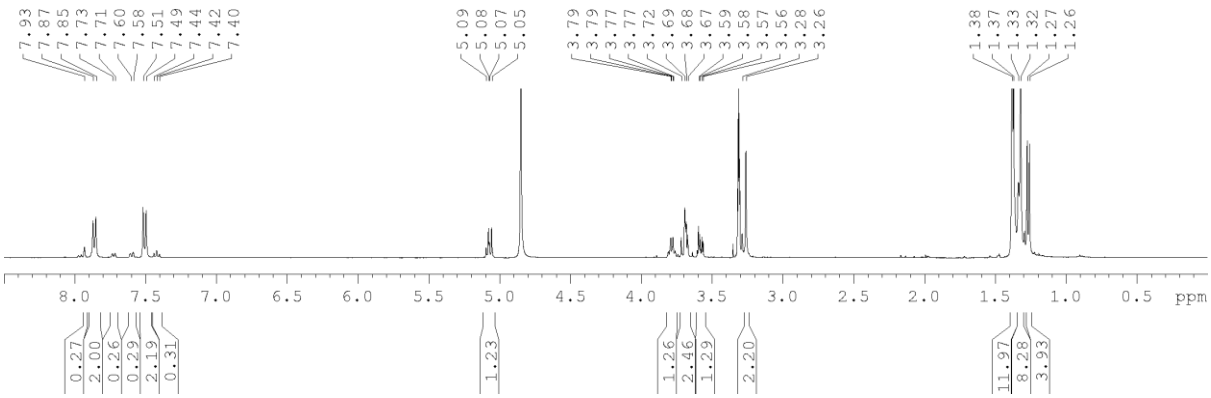
para-isomer:

δ = 7.86 (d, J = 8.4 Hz, 2H, *CH Ar*), 7.50 (d, J = 8.4 Hz, 2H, *CH Ar*), 5.06 (d, J_{1-2} = 9.0 Hz, 1H, *H-1*), 3.78 (dq, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 3.69 (mult., J_{2-1} = 9.2 Hz, J_{2-3} = 9.4 Hz, J_{4-3} =

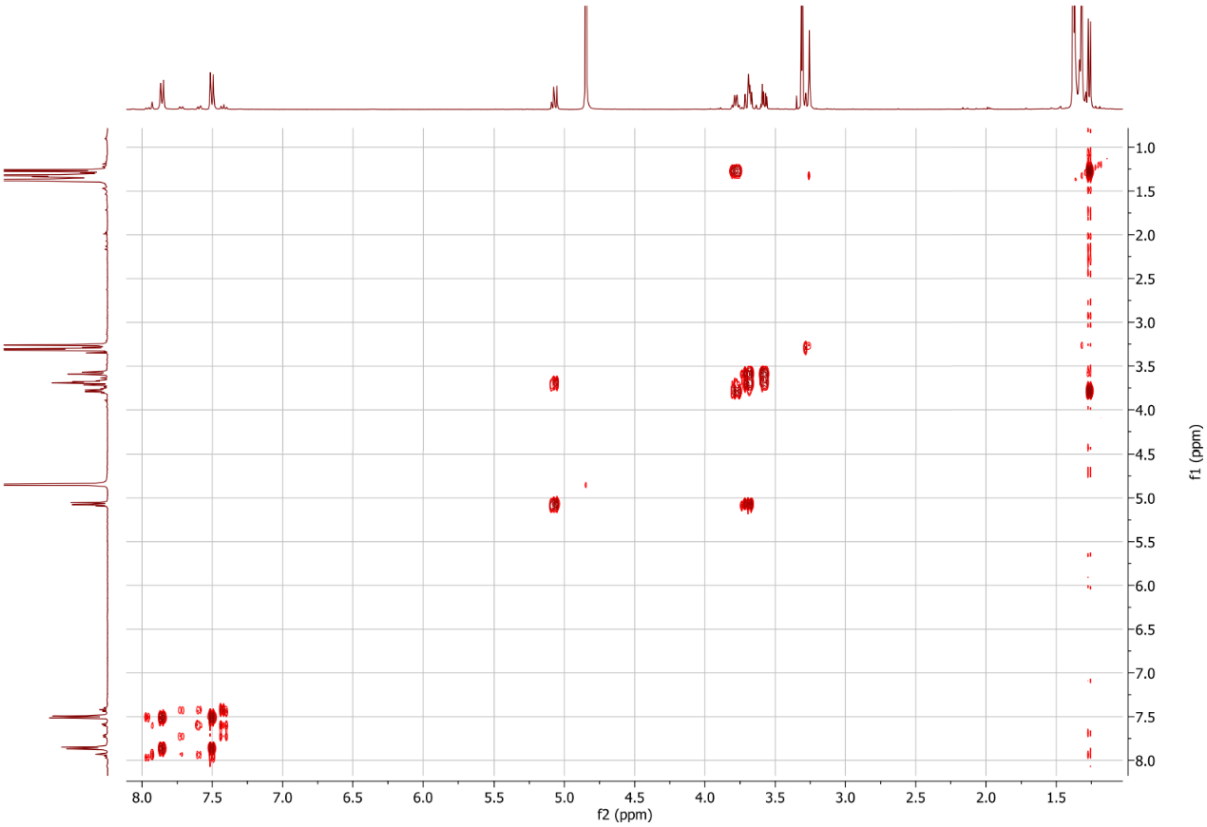
3.4 Hz, J_{4-5} = 1.1 Hz, 2H, *H-2* + *H-4*), 3.58 (dd, J_{3-2} = 9.5 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 3.26 (bs, 2H, CH_2), 1.38 (bs, 9H, *tBu*), 1.32 (bs, 6H, 2xCH_3), 1.26 (d, J_{CH3-5} = 6.5 Hz, 3H, *H-6*).

meta-isomer:

δ = 7.93 (bs, 1H, *CH Ar*), 7.72 (d, 1H, *Ar*), 7.59 (d, 1H, *Ar*), 7.42 (dd, H, *Ar*), 5.08 (d, J_{1-2} = 9.0 Hz, 1H, *H-1*), 3.78 (dq, 1H, *H-5*), 3.69 (mult., 2H, *H-2* + *H-4*), 3.58 (dd, 1H, *H-3*), 3.28 (bs, 2H, CH_2), 1.37 (bs, 9H, *tBu*), 1.33 (bs, 6H, 2xCH_3), 1.27 (d, 3H, *H-6*).



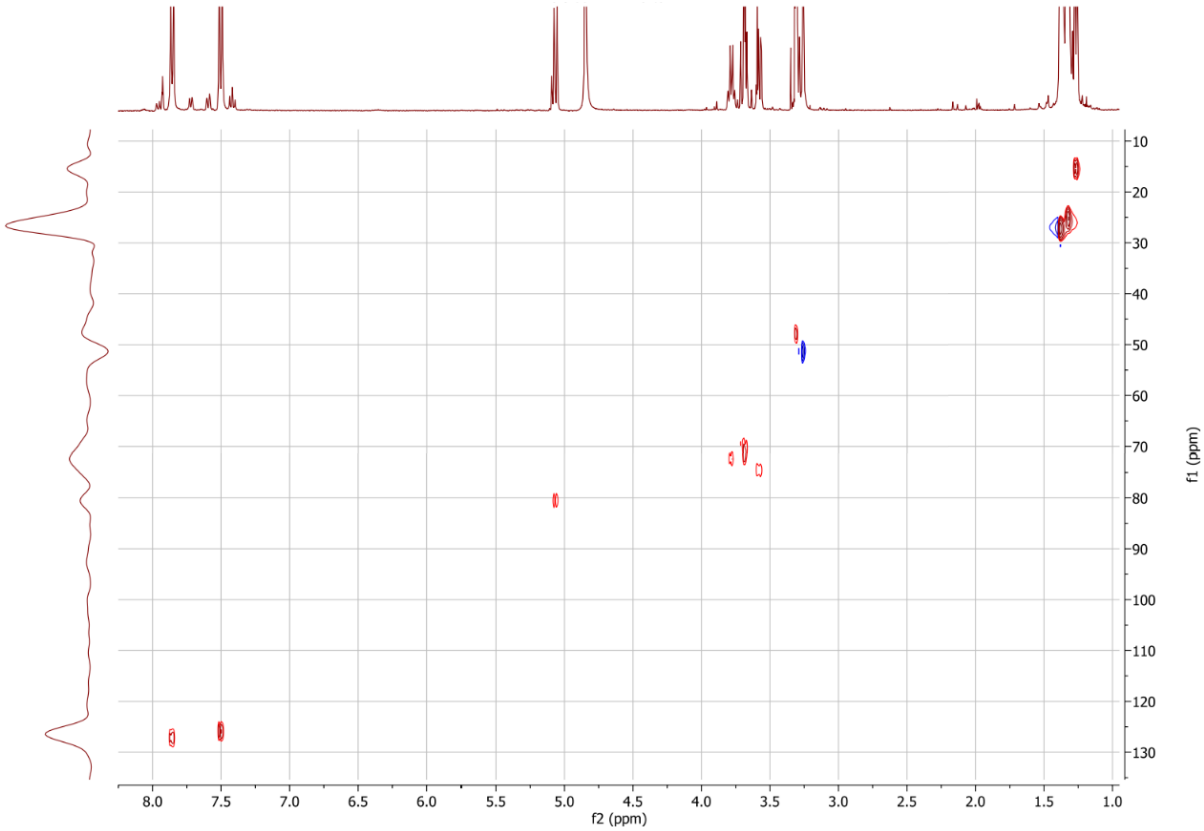
COSY:



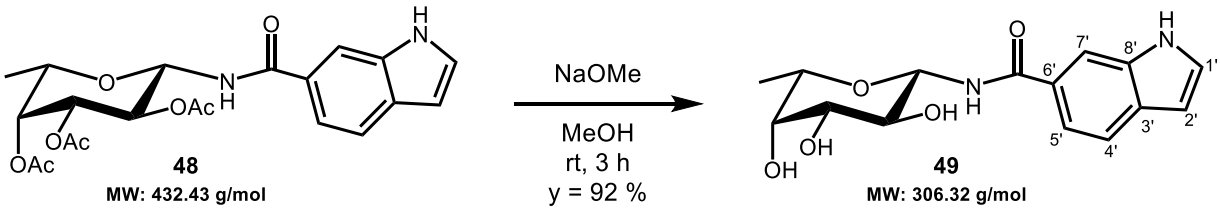
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 127.1, 125.9 (*para CH Ar*), 80.5 (*C1*), 74.5 (*C3*), 72.3 (*C5*), 71.5 (*C4*), 69.4 (*C2*), 51.5 (CH_2), 27.2 (CH_3 *tBu*), 25. (2xCH_3), 15.3 (*C6*).

HSQC:



(49) Synthesis and characterization of **N-(β-L-fucopyranosyl)-1H-indole-6-carboxamide (49)**:

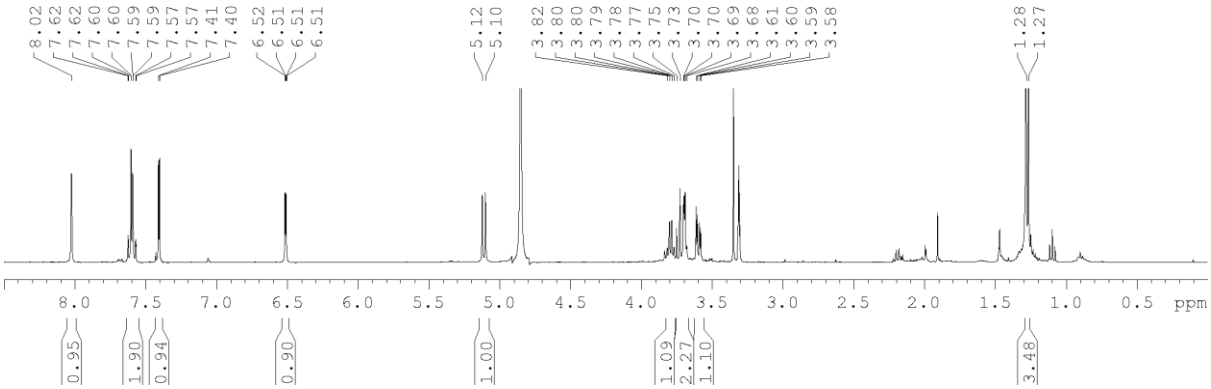


N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-indole-6-carboxamide 48 (0.046 mmol) was subjected to the aforementioned procedure to afford **49** (0.042 mmol, y = 92 %). TLC R_f (DCM/Acetone: 1/1): 0.05. $[\alpha]_D^{24.3}$ = -20.0 (MeOH, c 0.5). MS (ESI) calculated for $C_{15}H_{18}N_2O_5$ $[M + Na]^+$ *m/z*: 329.11; found: 329.09. HRMS (ESI^+-TOF) *m/z*: calculated for $C_{15}H_{18}N_2O_5$ $[M + H]^+$: 307.1289, found: 307.1286.

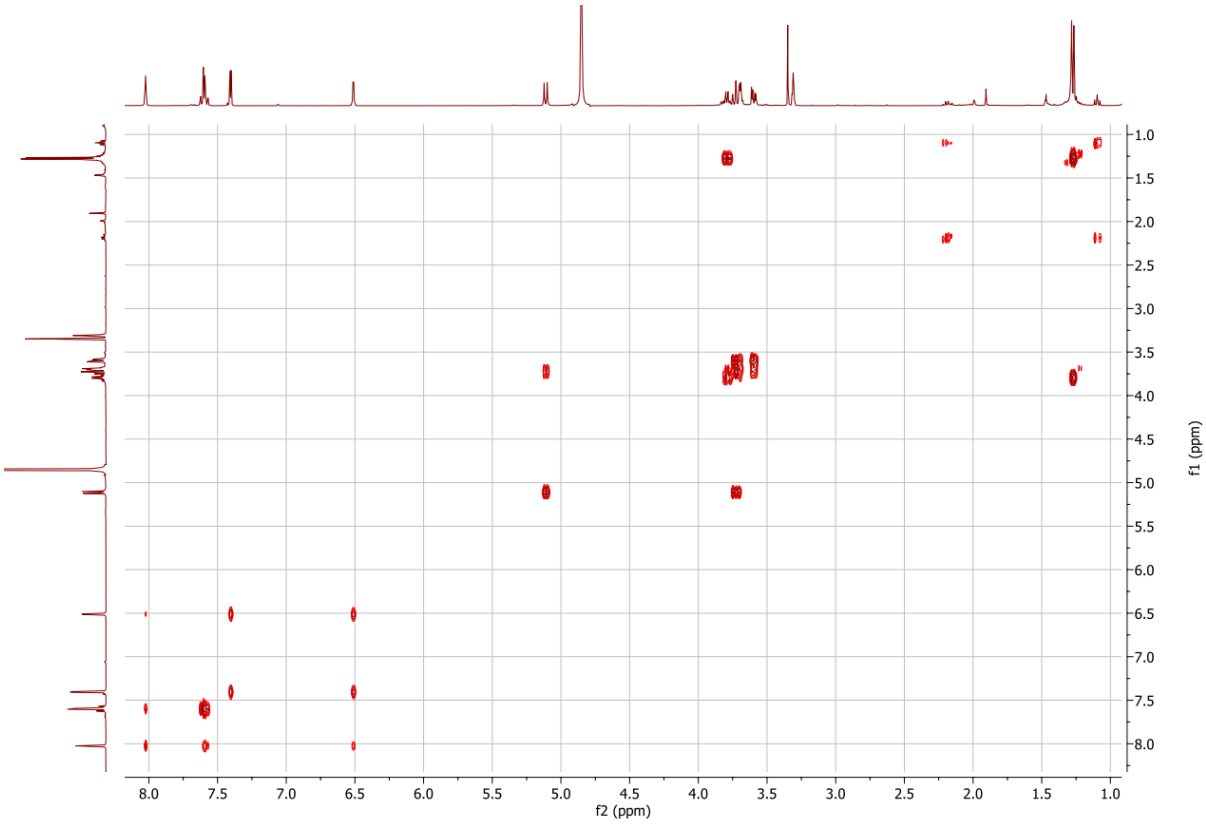
^{1}H NMR (400 MHz, MeOD):

δ = 8.02 (s, 1H, *H-4'*), 7.64-7.56 (mult., 2H, *H-5'*+ *H-7'*), 7.41 (d, $J_{3'-2'}$ = 3.1 Hz, 1H, *H-1'*), 6.51 (dd, $J_{2'-3'}$ = 3.1 Hz, $J_{2'-NH}$ = 0.8 Hz, 1H, *H-2'*), 5.11 (d, J_{1-2} = 9.0 Hz, 1H, *H-1*), 3.79 (m, 1H, *H-5*),

3.76-3.67 (mult., 2H, *H-2 + H-4*), 3.59 (dd, J_{3-2} = 9.5 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 1.27 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).

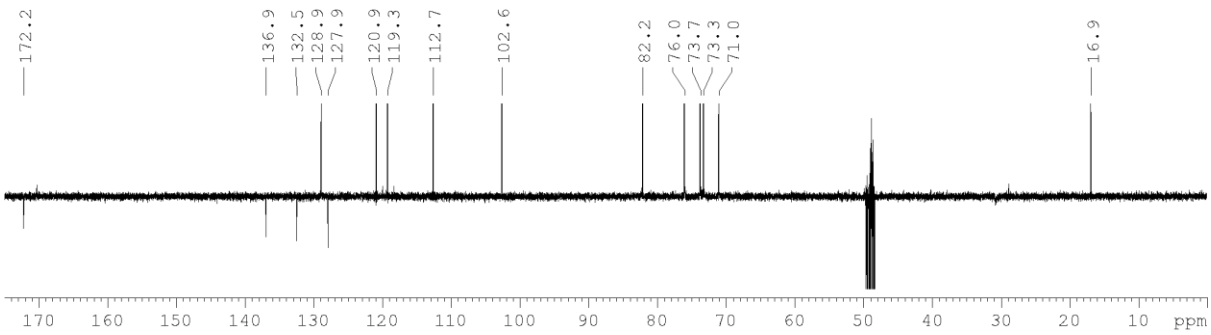


COSY:

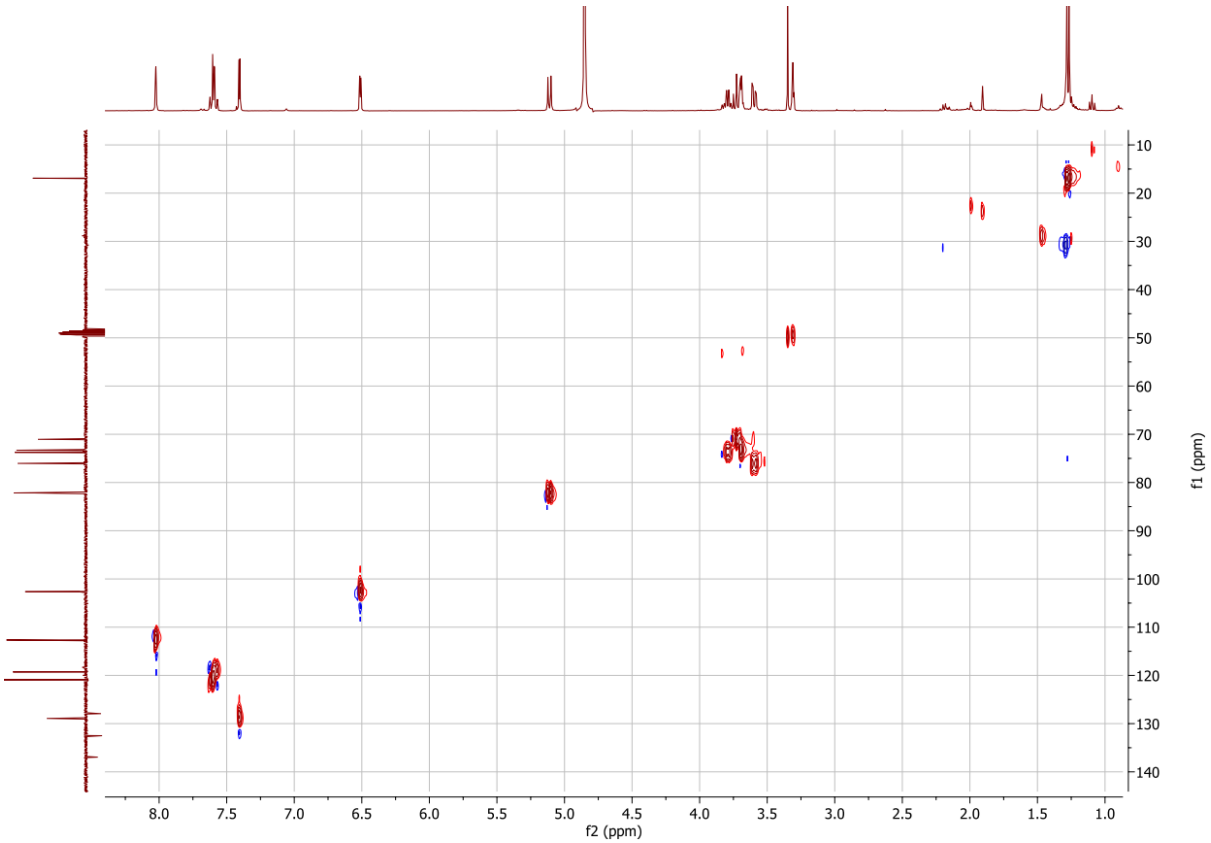


^{13}C NMR (400 MHz, MeOD):

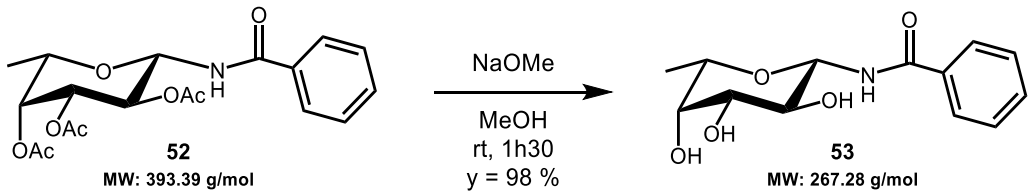
δ = 172.2 (*C=O*), 136.9, 132.5, 127.9 (*C hAr*), 128.9 (*C1'*), 120.9, 119.3 (*C5'* + *C7'*), 112.7 (*C4'*), 102.6 (*C2'*), 82.2 (*C1*), 76.0 (*C3*), 73.7 (*C5*), 73.3 (*C4*), 71.0 (*C2*), 17.0 (*C6*).



HSQC:



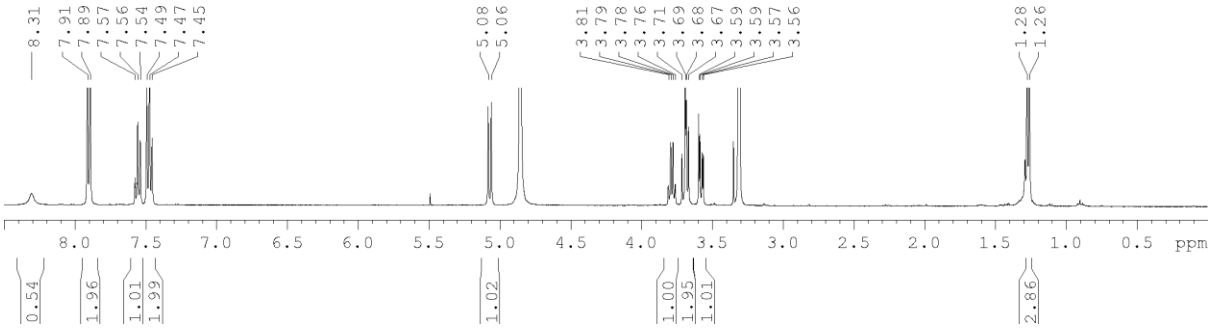
(53) Synthesis and characterization of **N-(β-L-fucopyranosyl)benzamide (53)**:



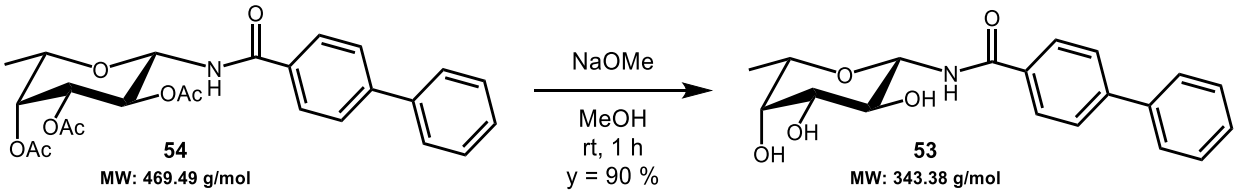
N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-benzamide 52 (0.076 mmol) was subjected to the aforementioned procedure to afford **53** (0.075 mmol, y = 98 %). TLC R_f (DCM/MeOH: 9/1): 0.31. $[\alpha]_D^{23.1}$ = -16.0 (MeOH, c 0.5). MS (ESI) calculated for $C_{13}H_{17}NO_5$ $[M + Na]^+$ *m/z*: 290.10; found: 290.03. HRMS (ESI^+-TOF) *m/z*: calculated for $C_{13}H_{17}NO_5$ $[M + H]^+$: 268.1180, found: 268.1179.

^{1}H NMR (400 MHz, MeOD):

δ = 8.31 (bs, 1H, *NH*), 7.90 (d, 2H, *CH Ar*), 7.56 (m, 1H, *CH Ar*) 7.46 (d, 2H, *CH Ar*), 5.07 (d, $J_{1\text{-}2}$= 9.0 Hz, 1H, *H-1*), 3.78 (dq, $J_{5\text{-}4}$ = 1.1 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.69 (mult., 2H, *H-2 + H-4*), 3.59 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$= 3.4 Hz, 1H, *H-3*), 1.27 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*).



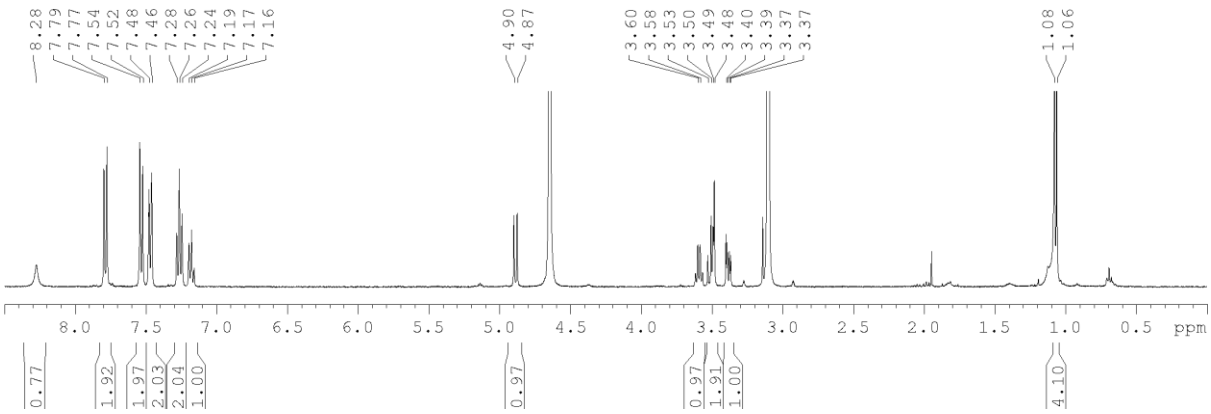
(55) Synthesis and characterization of **N-(β-L-fucopyranosyl)-[1,1'-biphenyl]-4-carboxamide (55)**:



N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-[1,1'-biphenyl]-4-carboxamide 54 (0.042 mmol) was subjected to the aforementioned procedure to afford **55** (0.038 mmol, y = 90 %). TLC R_f (DCM/MeOH: 95/5): 0.14.

^{1}H NMR (400 MHz, MeOD):

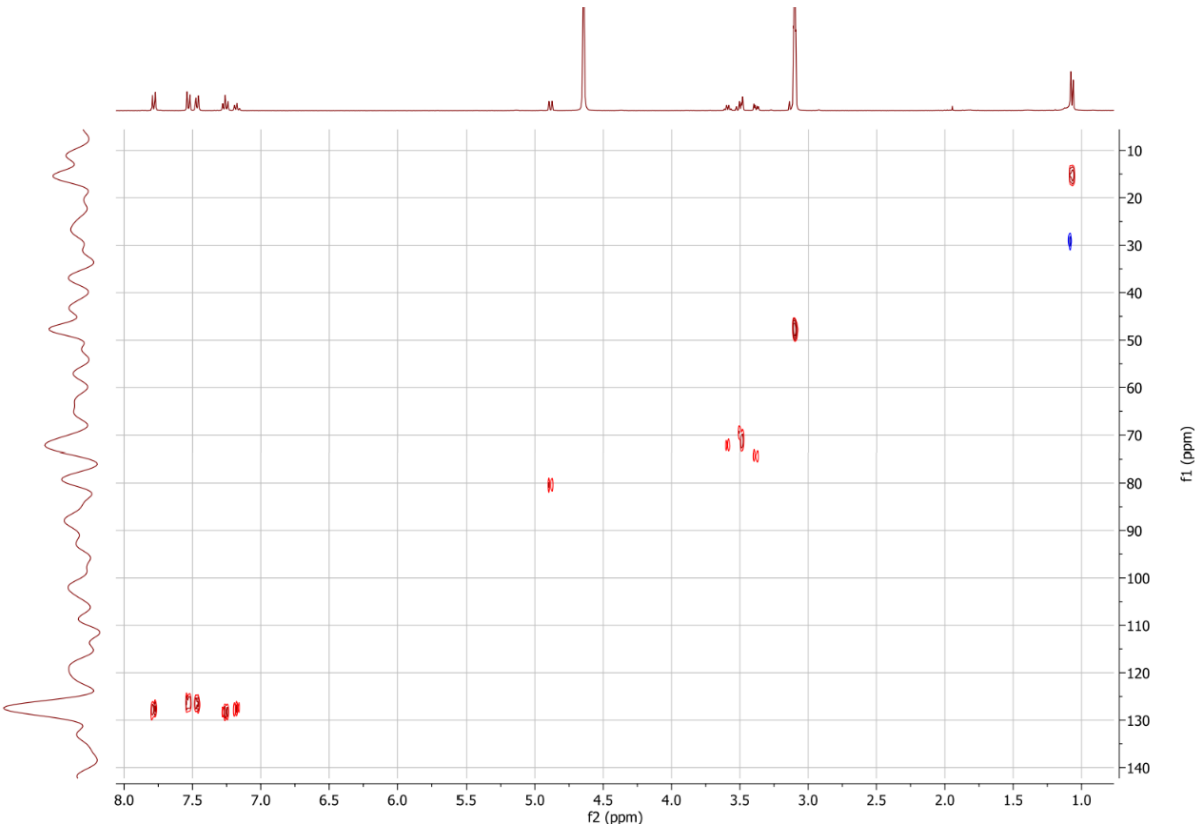
δ = 8.28 (bs, 1H, *NH*), 7.78 (d, 2H, *CH Ar*), 7.53 (d, 2H, *CH Ar*) 7.47 (d, 2H, *CH Ar*), 7.67 (t, 2H, *CH Ar*), 7.17 (m, 1H, *CH Ar*), 4.89 (d, $J_{1\text{-}2}$= 9.0 Hz, 1H, *H-1*), 3.59 (dq, $J_{5\text{-}4}$ = 1.0 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.50 (mult., 2H, *H-2 + H-4*), 3.38 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$= 3.4 Hz, 1H, *H-3*), 1.07 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*).



^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 128.2, 127.6, 127.5, 126.5, 126.4 (*CH Ar*), 120.7 (*CH hAr*), 80.4 (C*1*), 74.3 (*C3*), 72.1 (*C5*), 71.5 (*C4*), 69.5 (*C2*), 15.2 (*C6*).

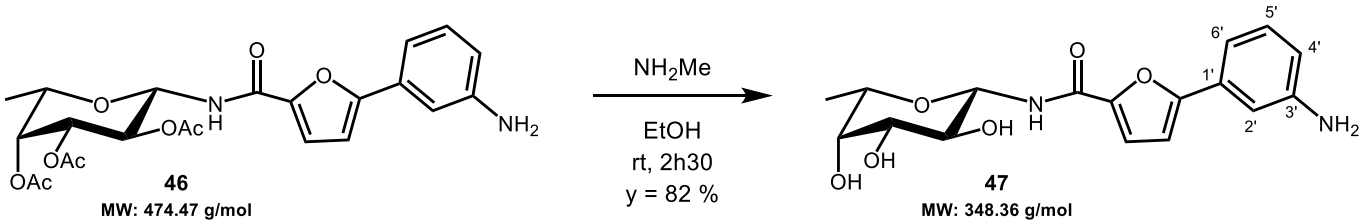
HSQC:



General procedure for deacetylation with NH_2Me, from Hribernik and co-workers:[233]

To the **acetylated compound** (1.0 eq) dissolved in EtOH (concentration: 0.05 M) was added a 8M solution of NH_2Me in EtOH (final concentration: 4 M). The reaction mixture stirred at room temperature for 2 - 3 h until TLC showed completion, before being concentrated or lyophilized to remove all by-products. The compound was used directly without further purification.

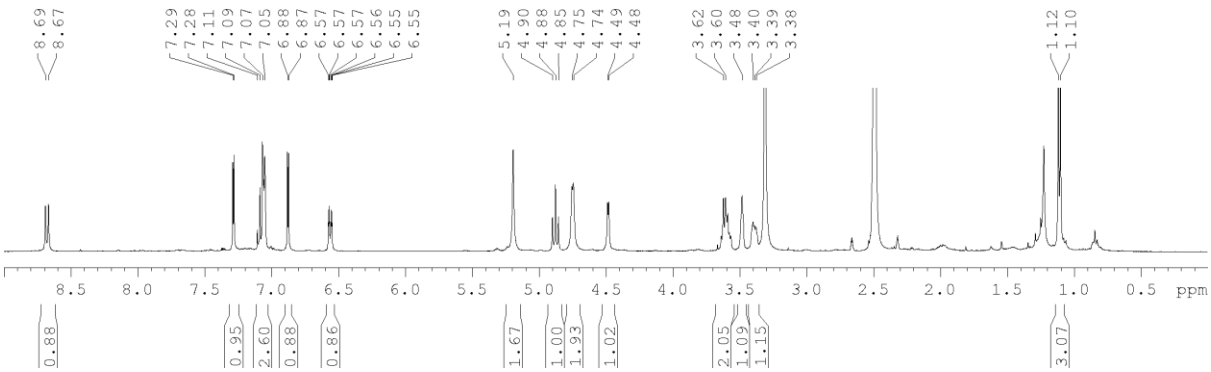
(47) Synthesis and characterization of **5-(3-aminophenyl)furan-2-carboxamido-(β-L-fucopyranose) (47)**:



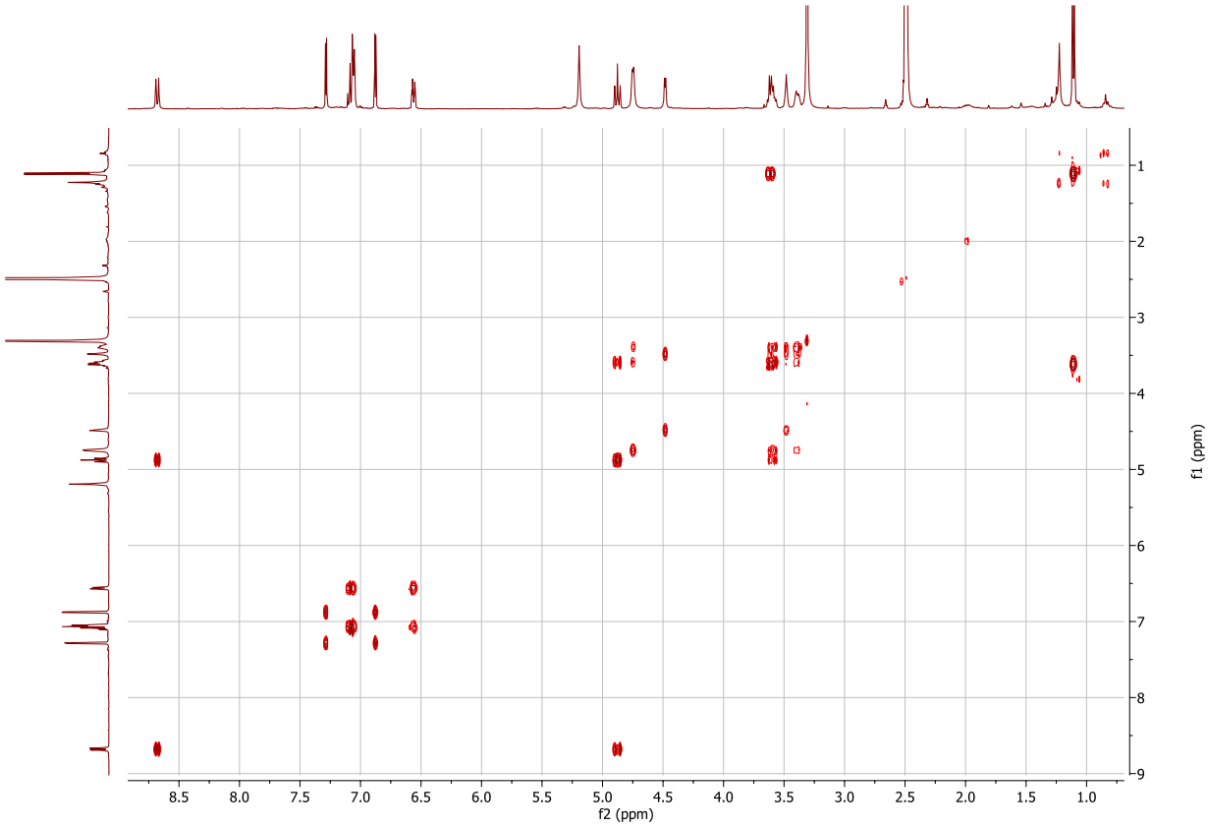
5-(3-aminophenyl)furan-2-carboxamido-(2,3,4-tri-O-acetyl-β-L-fucopyranose) 46 (0.017 mmol) was subjected to the aforementioned procedure to afford **47** (0.014 mmol, y = 82 %). TLC R_f (nHex/EtOAc: 2/8): 0.05. MS (ESI) calculated for $C_{17}H_{20}N_2O_6$ $[M + Na]^+$ *m/z*: 371.12; found: 371.15. HRMS (ESI^+-TOF) *m/z*: calculated for $C_{17}H_{20}N_2O_6$ $[M + Na]^+$: 371.1219, found: 371.1212.

^{1}H NMR (400 MHz, DMSO-d_6):

δ = 8.68 (d, J_{NH-1} = 9.1 Hz, 1H, *NH*), 7.28 (d, J = 3.6 Hz, 1H, *CH hAr*), 7.11 - 7.05 (mult., 3H, *H-2'* + *H-5'* + *H-6'*), 6.88 (d, J = 3.6 Hz, 1H, *CH hAr*), 6.56 (dt, J_{ortho} = 7.3 Hz, J_{meta} = 2.0 Hz, 1H, *H-4'*), 5.20 (s, 1H, *NH_2*), 4.88 (t, J_{NH-1} = J_{1-2}= 9.1 Hz, 1H, *H-1*), 4.75 (mult., 2H, *OH-2* + *OH-3*), 4.48 (d, J = 4.0 Hz, 1H, *OH-4*), 3.67 - 3.56 (mult., 2H, *H-2* + *H-5*), 3.48 (t, J_{3-4}= 3.5 Hz, 1H, *H-4*), 3.39 (m, 1H, *H-3*), 1.11 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*).

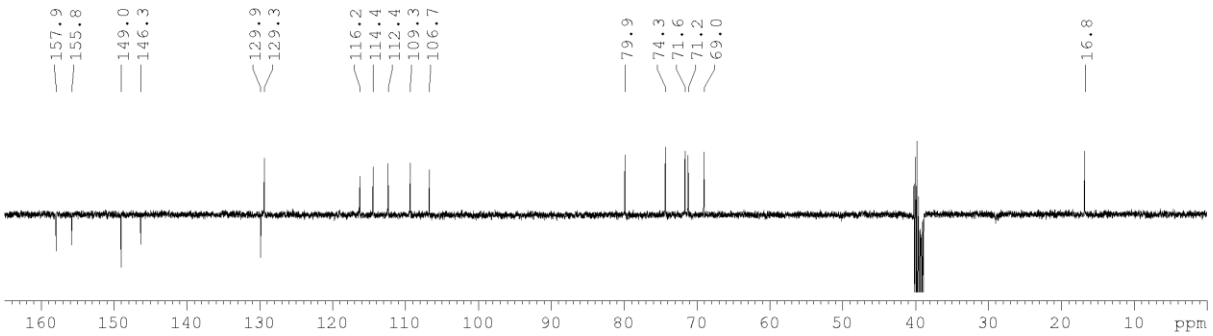


COSY:

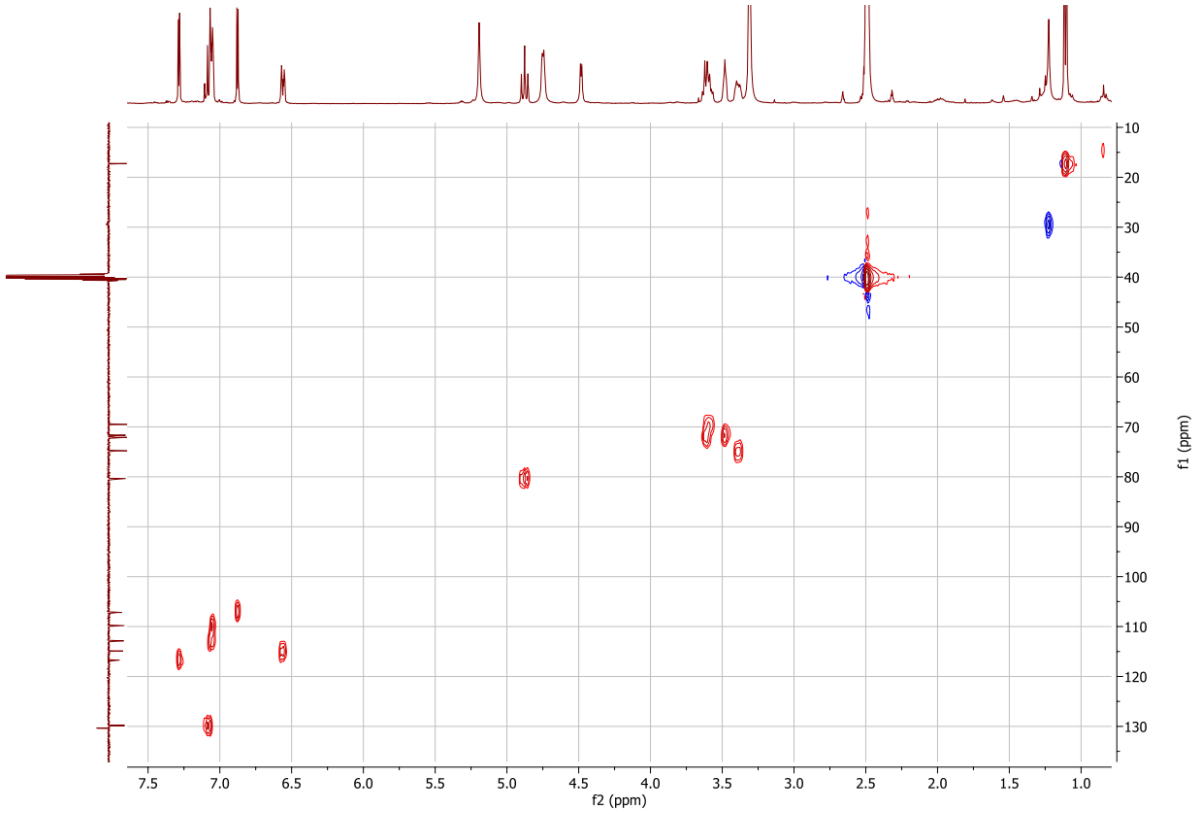


^{13}C NMR (400 MHz, DMSO-d$_6$):

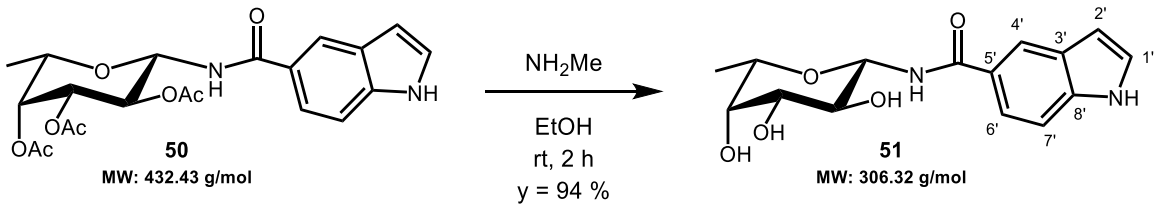
δ = 157.9 (*C=O*), 155.8 (*C$_{furane}$-Ar*), 149.0 (*C-NH$_2$*), 146.3 (*C$_{furane}$-C=O*), 129.9 (*C-C$_{furane}$*), 129.3 (*C5'*), 116.2 (*CH$_{furane}$*), 114.4 (*C4'*), 112.4, 109.3 (*C2' + C6'*), 106.7 (*CH$_{furane}$*), 79.9 (*C1*), 74.3 (*C3*), 71.6 (*C5*), 71.2 (*C4*), 69.0 (*C2*), 16.8 (*C6*).



HSQC:



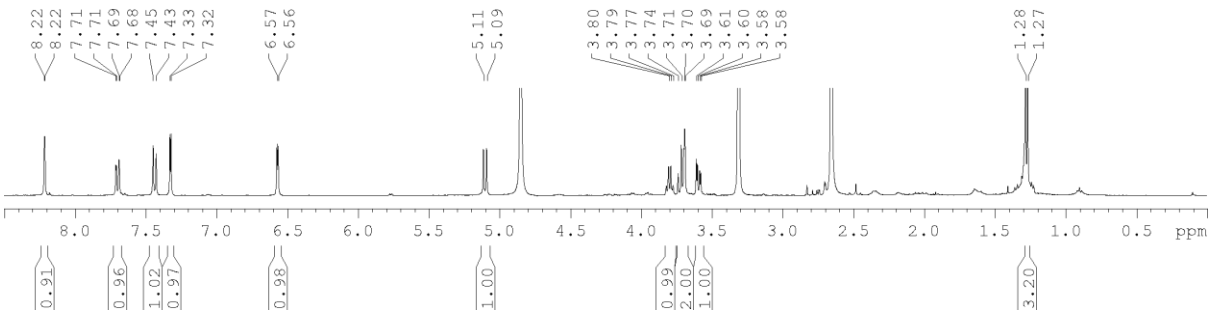
(51) Synthesis and characterization of **N-(β-L-fucopyranosyl)-1H-indole-5-carboxamide (51)**:



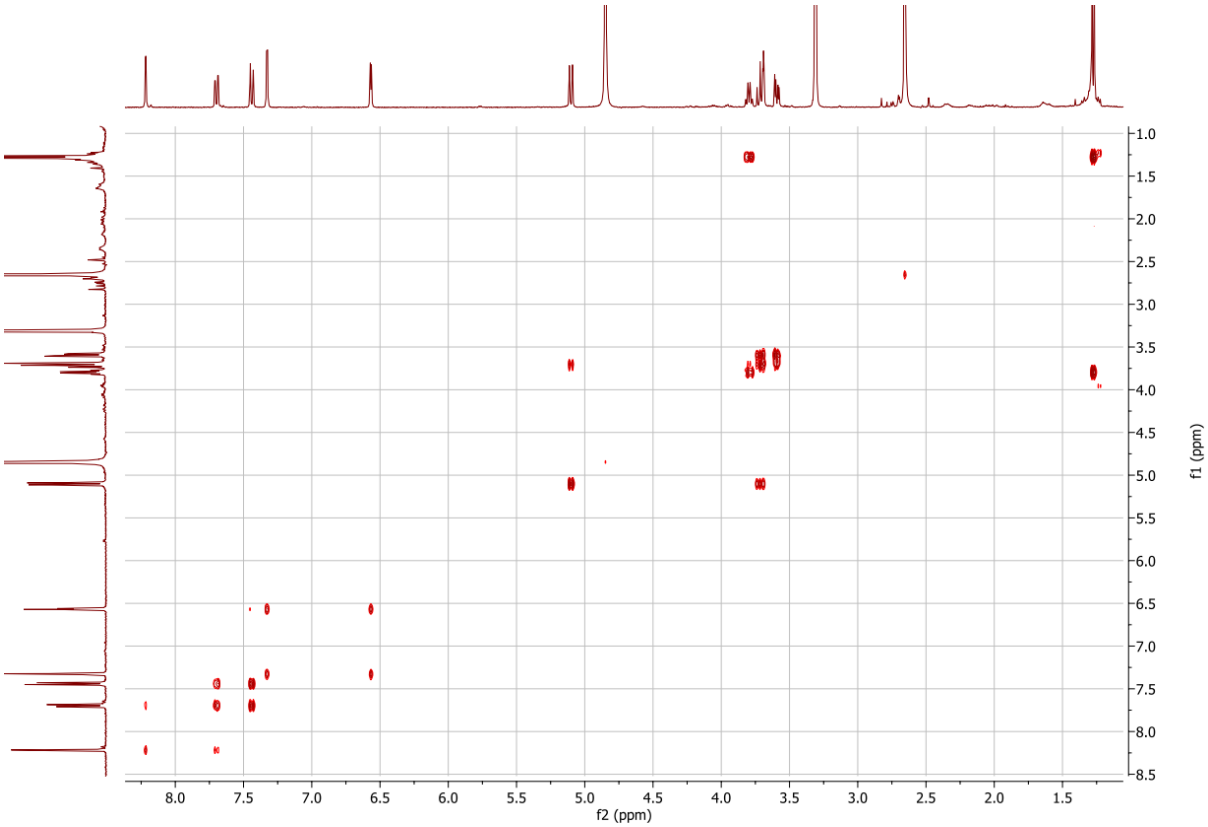
N-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-indole-5-carboxamide 50 (0.014 mmol) was subjected to the aforementioned procedure to afford **51** (0.013 mmol, y = 94 %). TLC R_f (DCM/Acetone: 1/1): 0.05. HRMS (ESI$^+$-TOF) *m/z*: calculated for $C_{15}H_{18}N_2O_5$ [M + H]$^+$: 307.1289, found: 307.1285.

^{1}H NMR (400 MHz, MeOD):

δ = 8.22 (d, J = 1.8 Hz, 1H, *H-4'*), 7.70 (dd, J' = 8.6 Hz, J = 1.8 Hz, 1H, *H-6'*), 7.44 (dd, J' = 8.6 Hz, J'' = 1.0 Hz, 1H, *H-7'*), 7.33 (d, J''' = 3.2 Hz, 1H, *H-1'*), 6.57 (dd, J''' = 3.2 Hz, J'' = 1.0 Hz, 1H, *H-2'*), 5.10 (d, $J_{1\text{-}2}$ = 9.1 Hz, 1H, *H-1*), 3.80 (dq, $J_{5\text{-}4}$ = 1.0 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.77-3.69 (mult., $J_{2\text{-}3}$ = 9.3 Hz, $J_{4\text{-}3}$ = 3.4 Hz, $J_{4\text{-}5}$ = 1.0 Hz, 2H, *H-2 + H-4*), 3.59 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.4 Hz, 1H, *H-3*), 1.27 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*).



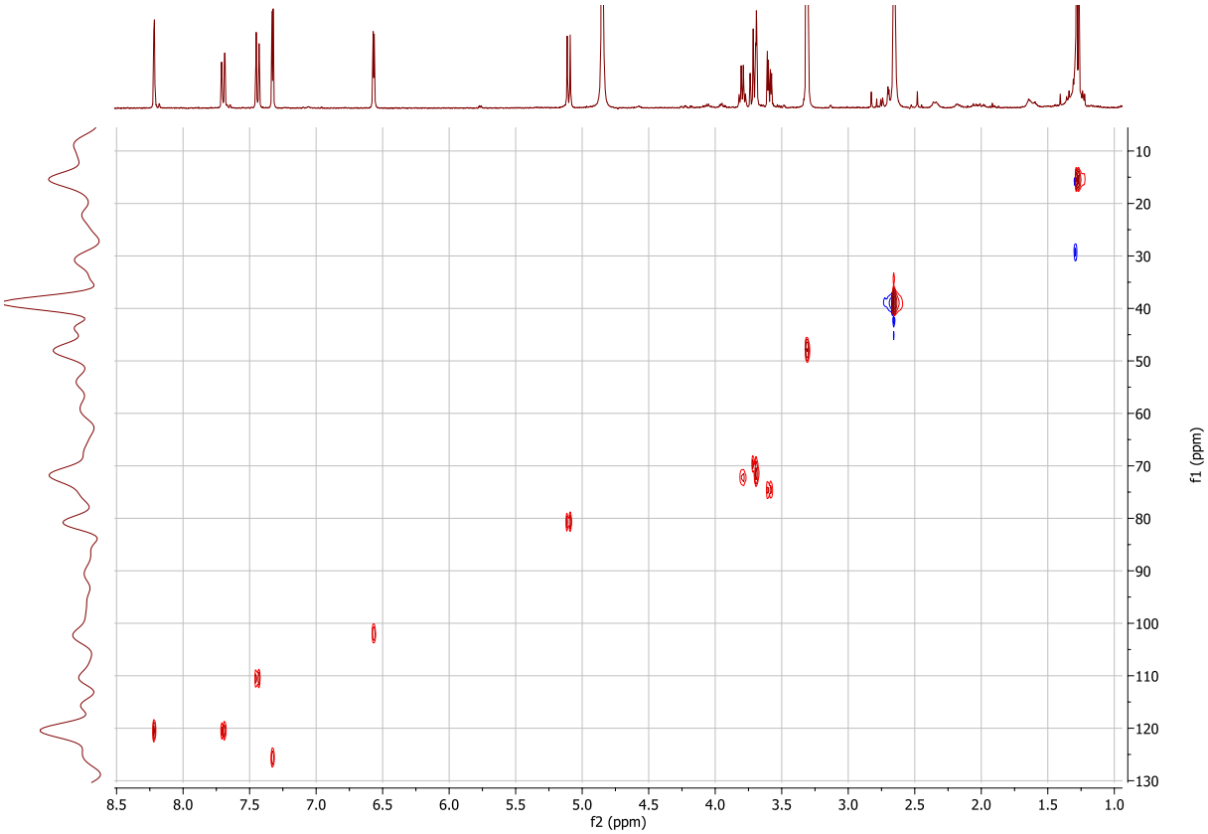
COSY:



^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 125.5 (*C1'*), 120.5 (*C4' + C6'*), 110.5 (*C7'*), 102.0 (*C2'*), 80.7 (*C1*), 74.6 (*C3*), 72.2 (*C5*), 71.5 (*C4*), 69.6 (*C2*), 15.5 (*C6*).

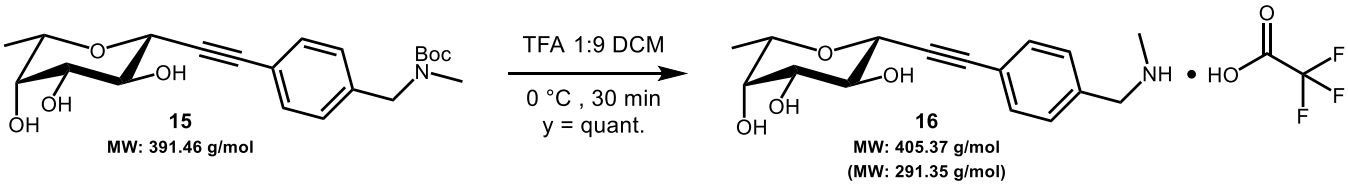
HSQC:


f2 (ppm)
f1 (ppm)
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
10
20
30
40
50
60
70
80
90
100
110
120
130

General procedure for Boc-removal, from Dedola and co-workers:[234]

The **Boc-protected compound** (1.0 eq) was dissolved in dry DCM (concentration: 10 mM) under N_2 atmosphere. The solution was cooled to 0 °C and TFA was added (ratio to DCM 1:9). The reaction mixture stirred while returning to room temperature for 0.5 - 1 h until TLC showed completion, before being diluted with toluene or MeOH and concentrated. The crude was used directly or after purification by automatic reverse phase chromatography (H_2O/CH_3CN).

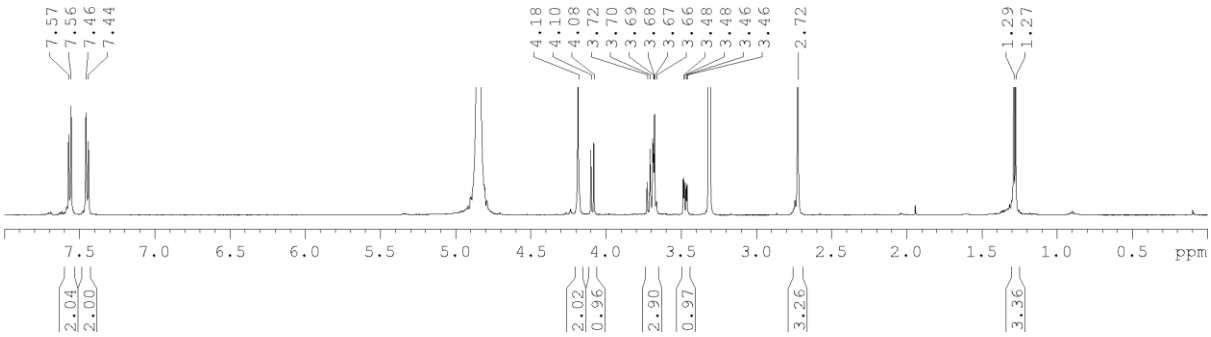
(16) Synthesis and characterization of **1-(4-(β-L-fucopyranosylethynyl)phenyl)-N-methylmethanamine (16)**:



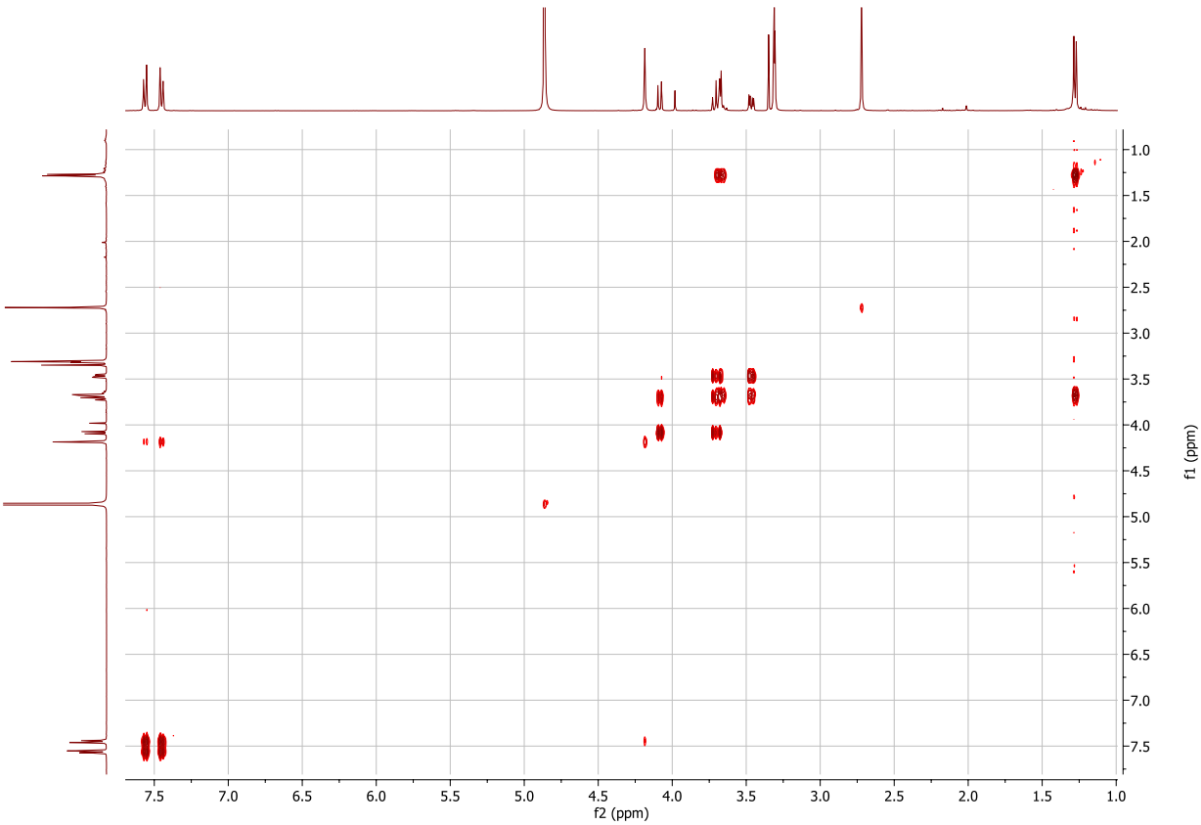
***tert*-butyl (4-(β-L-fucopyranosylethynyl)benzyl) (methyl) carbamate 15** (0.092 mmol) was subjected to the aforementioned procedure to afford **16** as a TFA salt (quantitative yield). The version of **16** that was a regioisomeric mixture was separated by HPLC: Gradient from 5 to 80 % of (CH_3CN/H_2O: 9/1; 0.1% TFA) in (H_2O; 0.1% TFA), with peaks coming out at 15 %. The *de novo* synthesis of the fragment circumvented the need for this separation during re-synthesis. TLC R_f (DCM/MeOH: 85/15): 0.20. $[\alpha]_D^{17}$ = 0.4 (MeOH, c 1). MS (ESI) calculated for $C_{16}H_{21}NO_4$ $[M + H]^+$ *m/z*: 292.15; found: 292.09. HRMS (ESI^+-TOF) *m/z*: calculated for $C_{16}H_{21}NO_4$ $[M + H]^+$: 292.1543, found: 292.1543.

^{1}H NMR (400 MHz, MeOD):

δ = 7.56 (d, J = 8.3 Hz, 2H, *CH Ar*), 7.45 (d, J = 8.3 Hz, 2H, *CH Ar*), 4.18 (bs, 2H, CH_2), 4.09 (d, J_{1-2} = 9.7 Hz, 1H, *H-1*), 3.69 (mult., $J_{2-1} = J_{2-3}$ = 9.7 Hz, 3H, *H-2* + *H-4* + *H-5*), 3.47 (dd, J_{3-2} = 9.5 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 2.72 (bs, 3H, *N-CH₃*), 1.28 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).

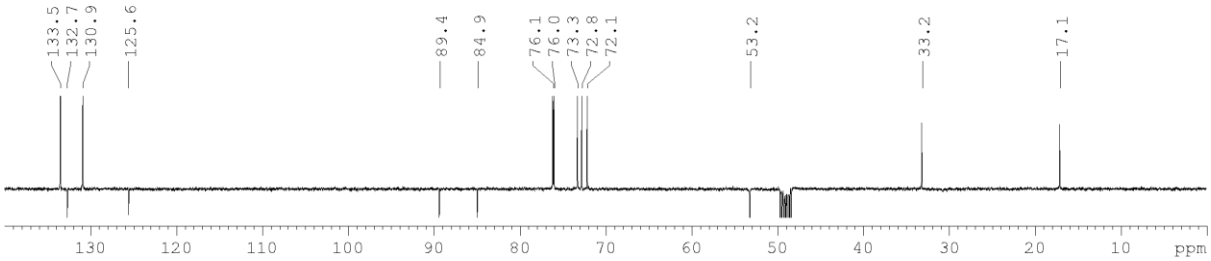


COSY:

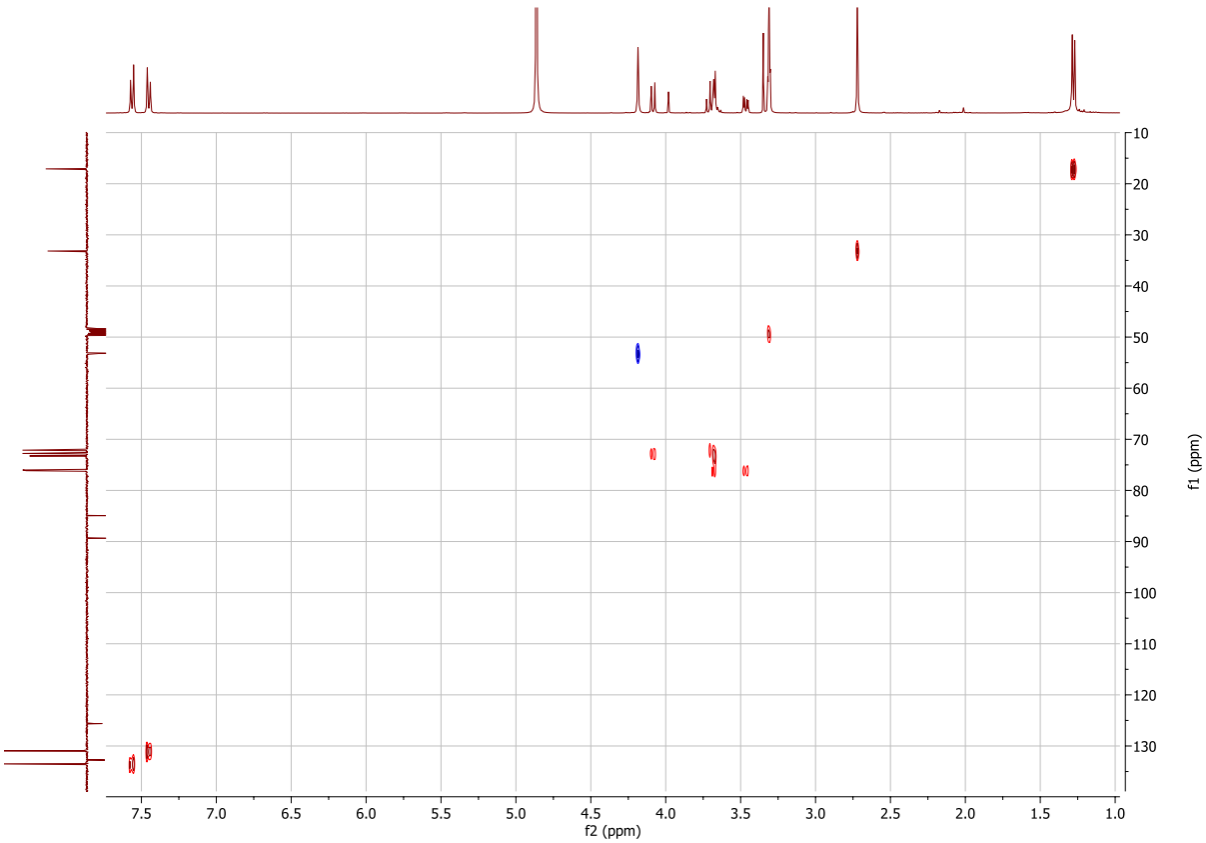


^{13}C NMR (400 MHz, MeOD):

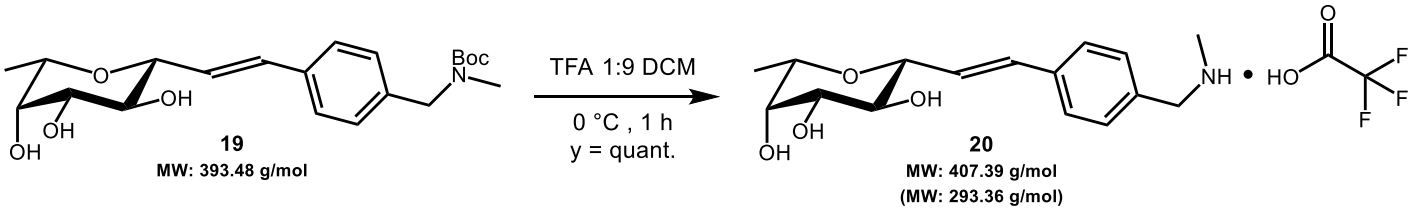
δ = 133.5, 130.9 (*CH Ar*), 132.7 (*CH₂-C Ar*), 125.6 (*≡C-C Ar*), 89.4 (*C1-C≡*), 84.9 (*≡C-Ar*), 76.1, 76.0 (*C3* + *C5*), 73.3 (*C2*), 72.8 (*C1*), 72.1 (*C4*), 53.2 (CH_2), 33.2 (CH_3), 17.1 (*C6*).



HSQC:



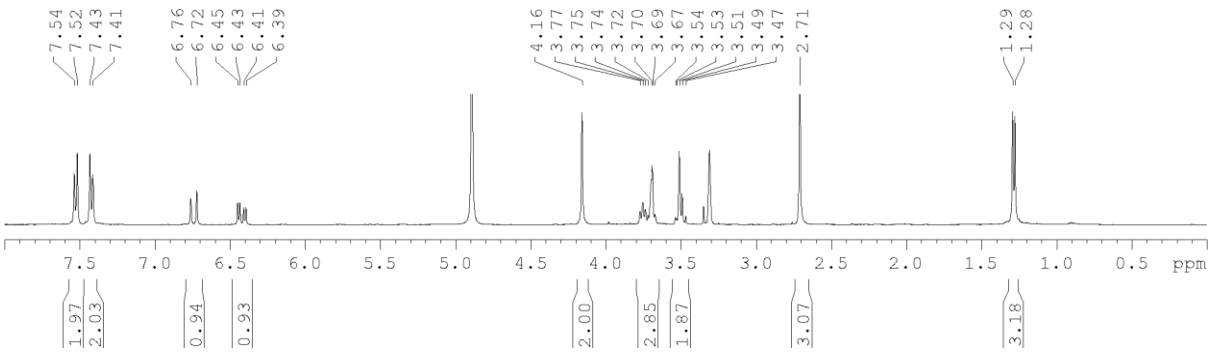
(20) Synthesis and characterization of **1-(4-(β-L-fucopyranosylvinyl)phenyl)-N-methylmethanamine (20)**:



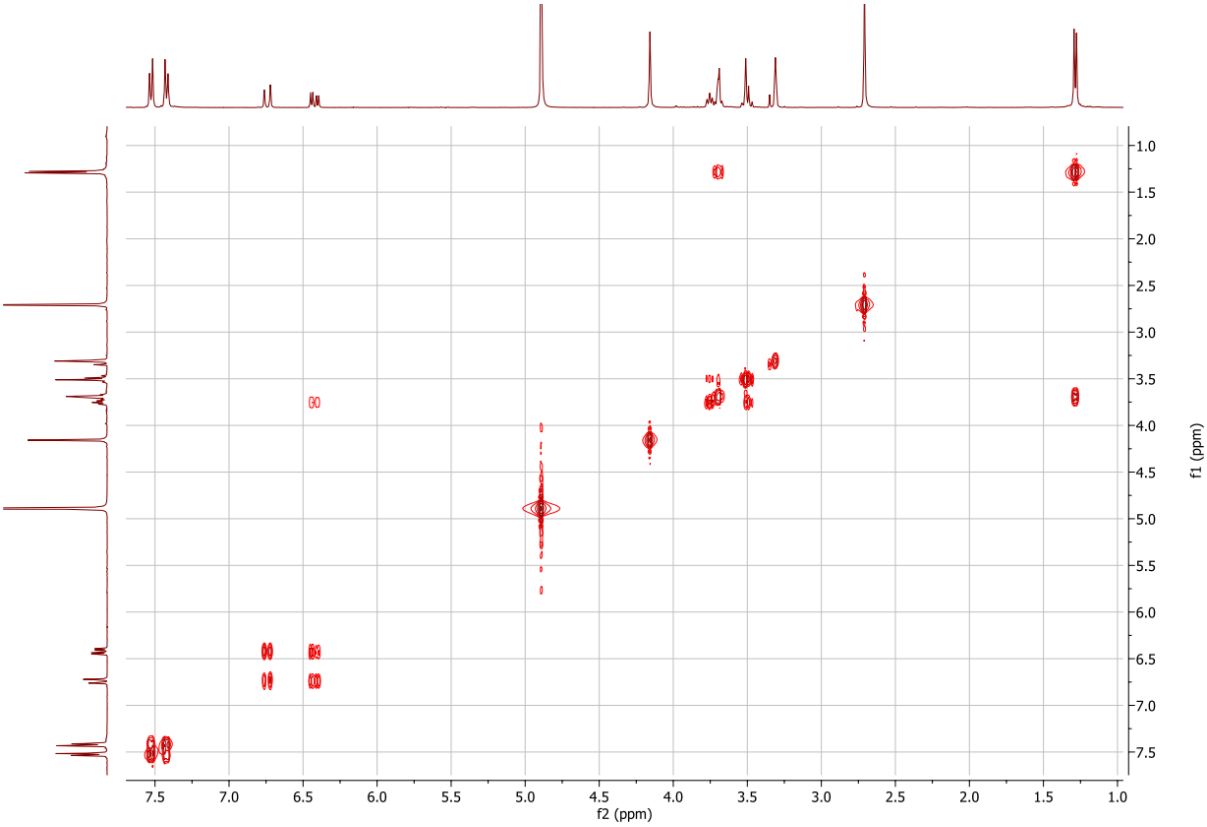
***tert*-butyl (*E*)-(4-(β-L-fucopyranosylvinyl)benzyl) (methyl) carbamate 19** (0.033 mmol) was subjected to the aforementioned procedure to afford **20** as a TFA salt (quantitative yield). TLC R_f (DCM/MeOH: 85/15): 0.20. $[\alpha]_D^{16.1}$ = 0.7 (MeOH, c 1). Exact Mass calculated for $C_{16}H_{23}NO_4$ $[M]^+$: 293.1627, found: 293.1xxx.

^{1}H NMR (400 MHz, MeOD):

δ = 7.53 (d, J = 8.2 Hz, 2H, *CH Ar*), 7.42 (d, J = 8.2 Hz, 2H, *CH Ar*), 6.74 (d, J_{trans} = 16.0 Hz, 1H, =*CH*-*Ar*), 6.42 (dd, J_{trans} = 16.0 Hz, J_{CH-1} = 6.2 Hz, 1H, *C1*-*CH*=), 4.16 (bs, 2H, CH_2), 3.77 - 3.67 (mult., J_{5-4} = 1.0 Hz, J_{5-CH3} = 6.5 Hz, 3H, *H-1* + *H-3* + *H-5*), 3.54 - 3.47 (mult., J_{2-1} = 9.4 Hz, 2H, *H-2* + *H-4*), 2.71 (bs, 3H, *N-CH₃*), 1.29 (d, J_{CH3-5} = 6.5 Hz, 3H, CH_3).

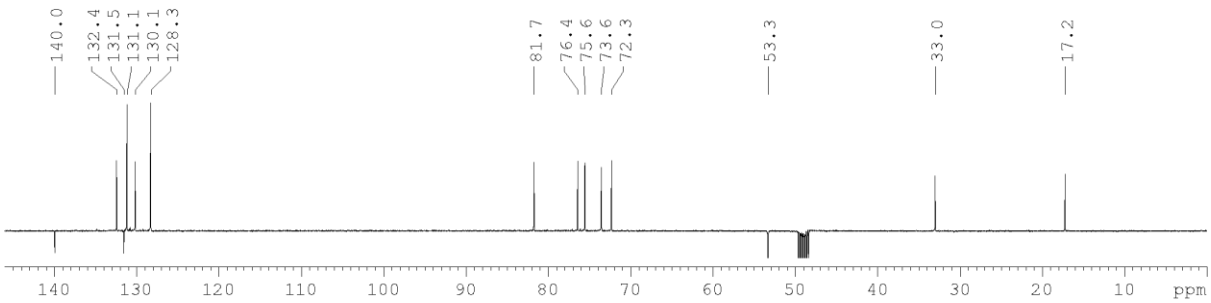


COSY:

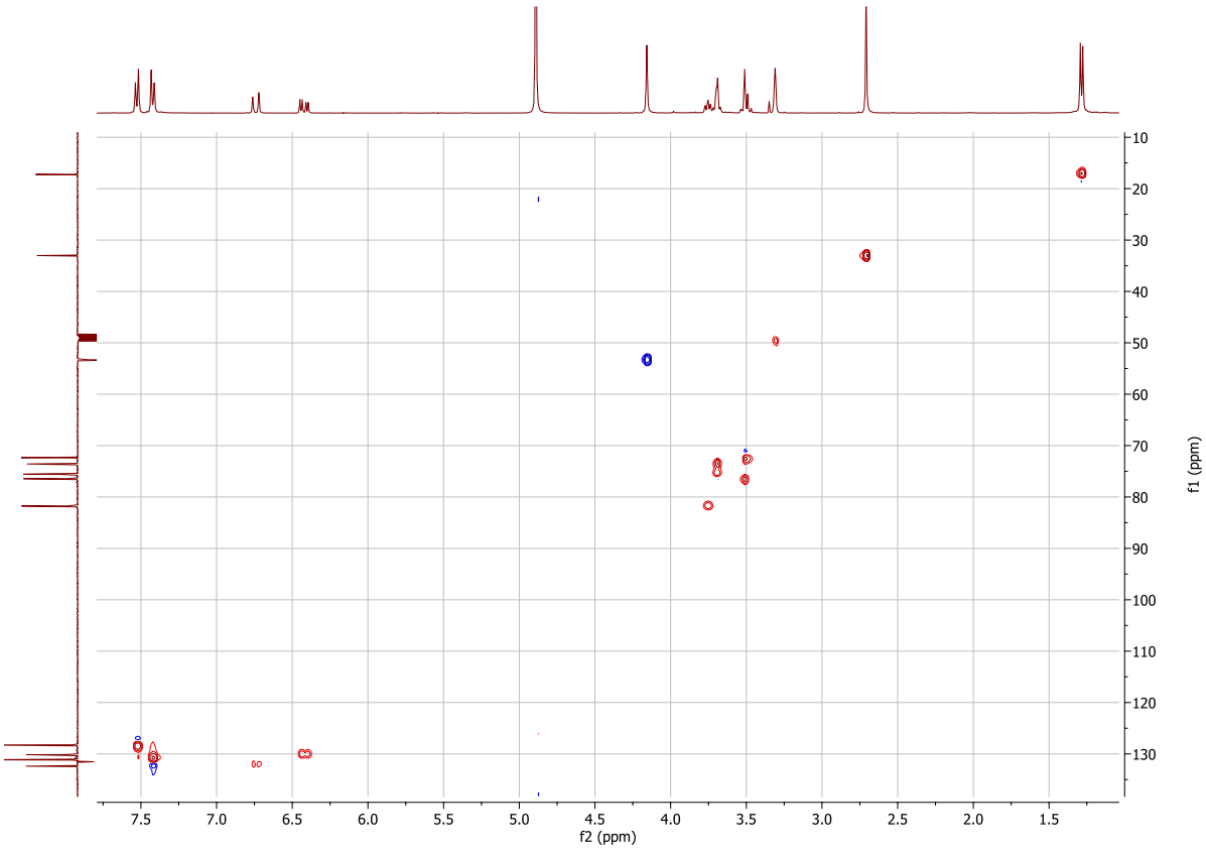


^{13}C NMR (400 MHz, MeOD):

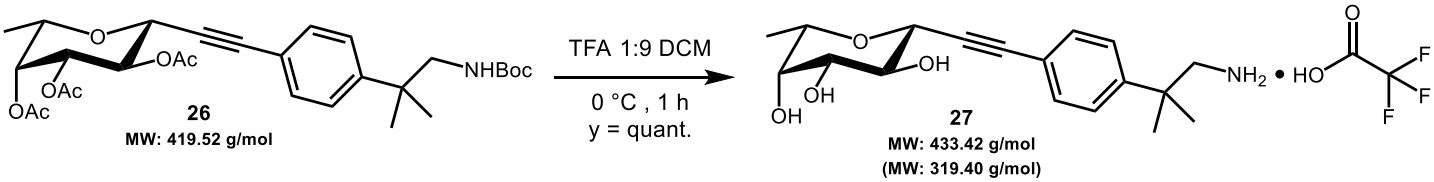
δ = 140.0 (*C Ar*), 132.4 (=<u>*CH*</u>-*Ar*), 131.5 (*C Ar*), 131.1 (*CH Ar*), 130.1 (*C1*-<u>*CH*</u>=), 128.3 (*CH Ar*), 81.7 (*C5*), 76.4 (*C2*), 75.6 (*C3*), 73.6 (*C1*), 72.3 (*C4*), 53.3 (CH_2), 33.0 (CH_3), 17.2 (*C6*).



HSQC:



(27) Synthesis and characterization of **(2-(4-(β-L-fucopyranosylethynyl)phenyl)-2-methylpropan-1-amine (27)**:



***tert*-butyl (2-(4-(β-L-fucopyranosylethynyl)phenyl)-2-methylpropyl)carbamate 26** (0.068 mmol) was subjected to the aforementioned procedure to afford regioisomeric *para*/*meta* mixture (75/25) **27** as a TFA salt (quantitative yield). TLC R_f (DCM/MeOH: 85/15): 0.30. MS (ESI) calculated for $C_{18}H_{25}NO_4$ $[M + H]^+$ *m/z*: 320.19; found: 320.12. HRMS (ESI^+-TOF) *m/z*: calculated for $C_{18}H_{25}NO_4$ $[M + H]^+$: 320.1856, found: 320.1852.

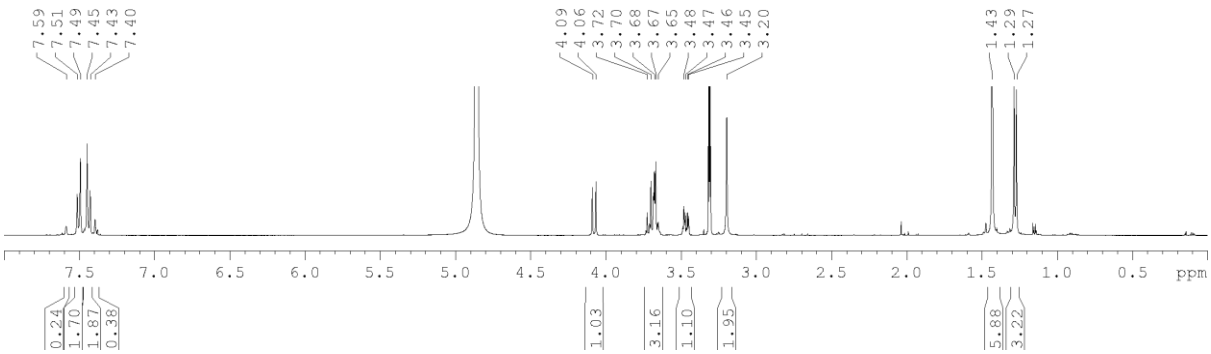
^{1}H NMR (400 MHz, MeOD): regioisomeric *para*/*meta* mixture (75/25)

para-isomer:

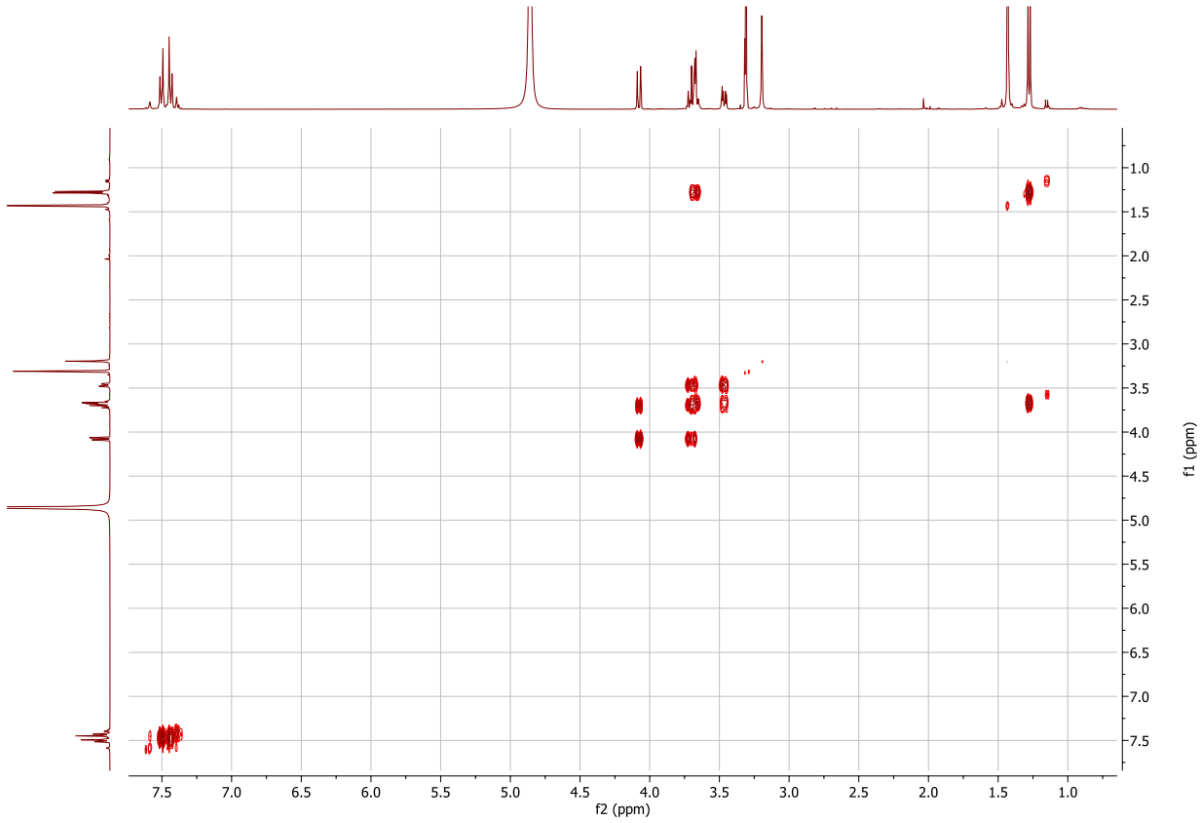
δ = 7.50 (d, J = 8.5 Hz, 2H, *CH Ar*), 7.44 (d, J = 8.5 Hz, 2H, *CH Ar*), 4.08 (d, $J_{1\text{-}2}$ = 9.6 Hz, 1H, *H-1*), 3.70 (mult., $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.5 Hz, 3H, *H-2* + *H-4* + *H-5*), 3.46 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.2 Hz, 1H, *H-3*), 3.20 (bs, 2H, CH_2), 1.43 (bs, 6H, $2xCH_3$), 1.28 (d, $J_{6\text{-}5}$ = 6.5 Hz, 6H, *H-6*).

meta-isomer:

δ = 7.59 (bs, 1H, *CH Ar*), 7.50 (m, 1H, *CH Ar*), 7.44 (m, 1H, *CH Ar*), 7.40 (m, 1H, *CH Ar*), 4.08 (d, $J_{1\text{-}2}$ = 9.6 Hz, 1H, *H-1*), 3.72 (mult., $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.5 Hz, 3H, *H-2* + *H-4* + *H-5*), 3.48 (dd, $J_{3\text{-}2}$ = 9.4 Hz, 1H, *H-3*), 3.20 (bs, 2H, CH_2), 1.43 (bs, 6H, $2xCH_3$), 1.28 (d, $J_{6\text{-}5}$ = 6.5 Hz, 6H, *H-6*).

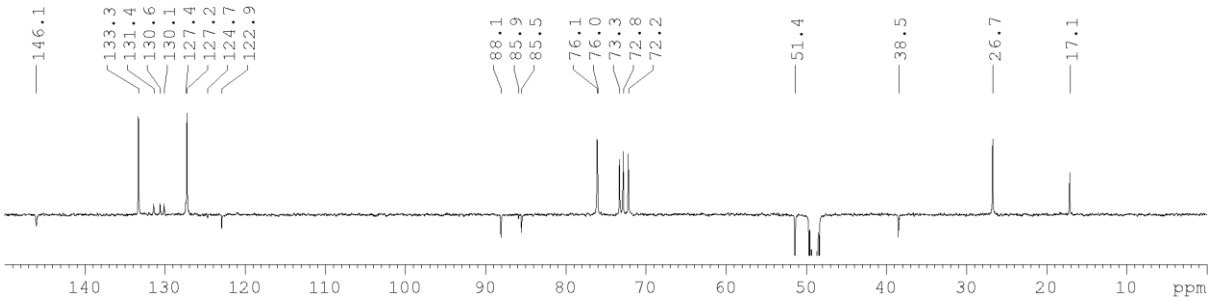


COSY:

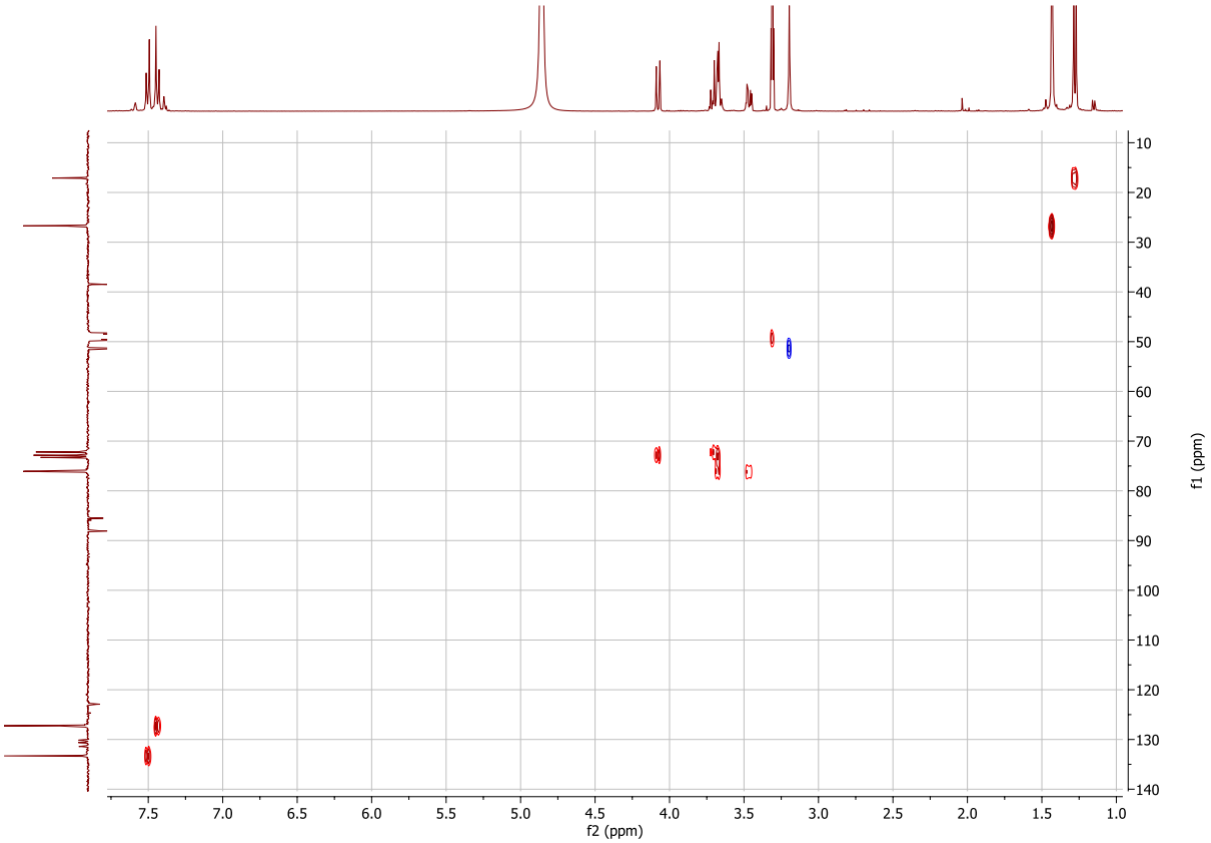


^{13}C NMR (400 MHz, MeOD):

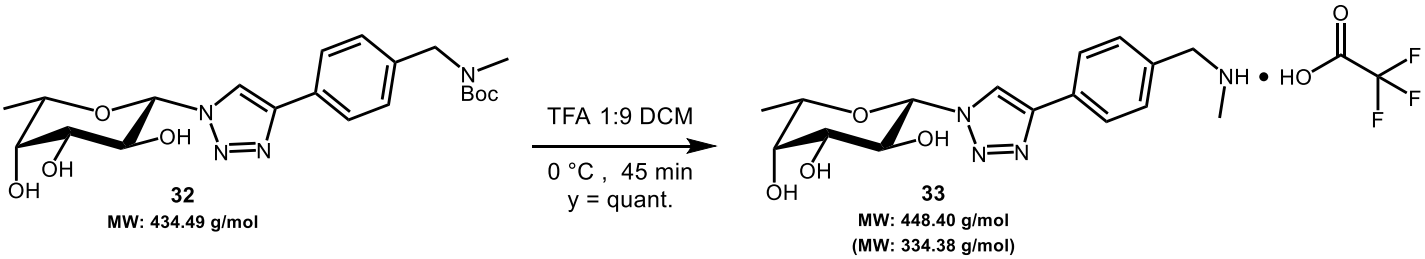
δ = 146.1 (*C(CH₃)₂-C Ar*), 133.3 (*para CH Ar*), 131.4, 130.6, 130.1, 127.4 (*meta CH Ar*), 127.3 (*para CH Ar*), 124.7 (*meta ≡C-C Ar*), 122.9 (*para ≡C-C Ar*), 88.1 (*C1-C≡*), 85.9 (*meta ≡C-Ar*), 85.5 (*para ≡C-Ar*), 76.1, 76.0 (*C3 + C5*), 73.3 (*C4*), 72.8 (*C1*), 72.2 (*C2*), 51.4 (CH_2), 38.5 ($C(CH_3)_2$), 26.7 (2xCH_3), 17.1 (*C6*).



HSQC:



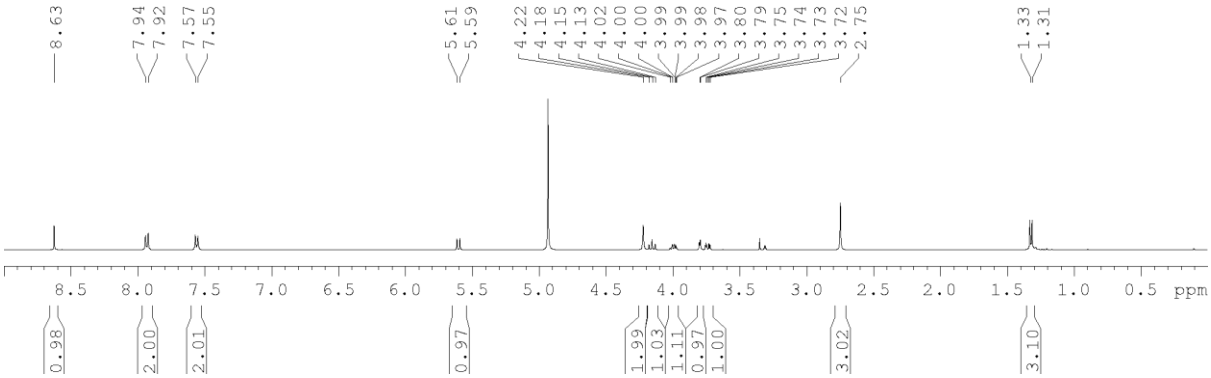
(33) Synthesis and characterization of **1-(4-(1-(β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)phenyl)-N-methylmethanamine (33)**:



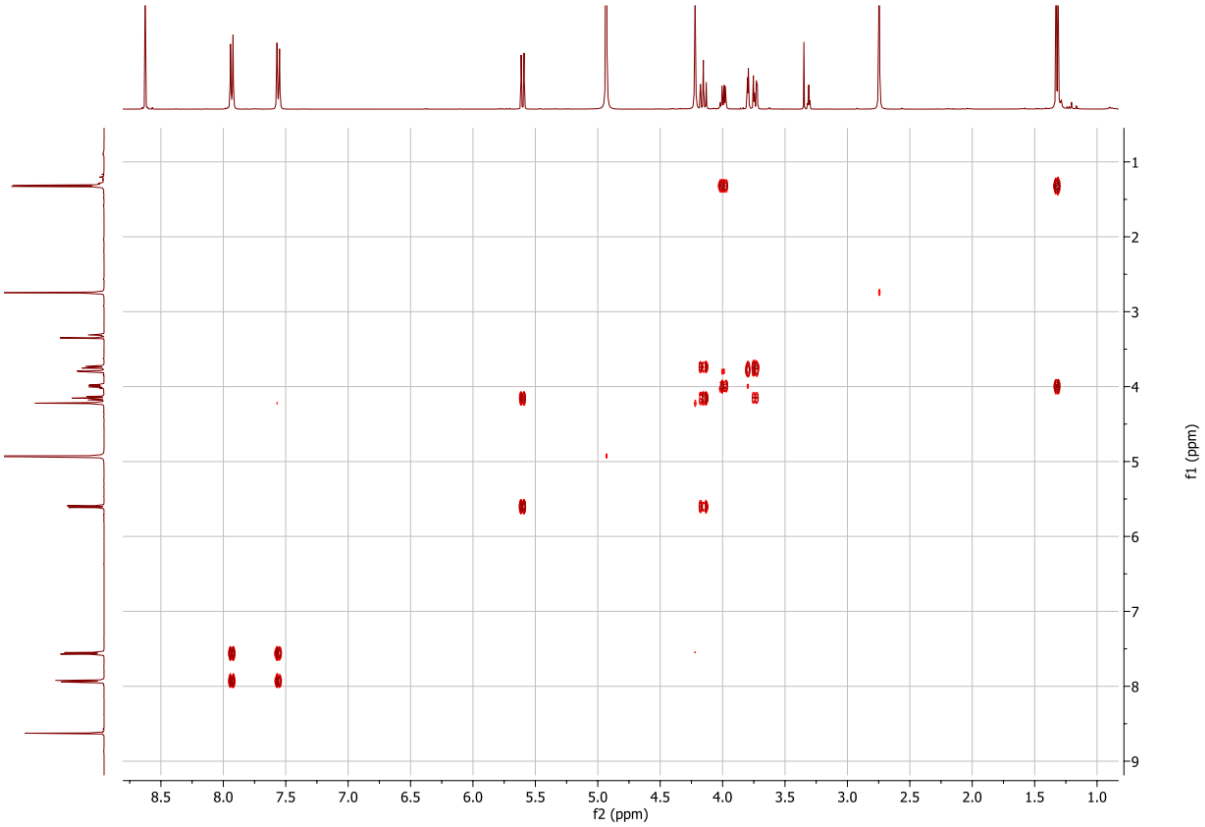
***tert*-butyl (4-(1-(2,3,4-tri-O-acetyl β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)benzyl) (methyl) carbamate 32** (0.143 mmol) was subjected to the aforementioned procedure to afford **33** as a TFA salt (quantitative yield). TLC R_f (DCM/MeOH: 75/25): 0.35. $[\alpha]_D^{17.6}$ = 5.5 (MeOH, c 1). Exact Mass calculated for $C_{16}H_{22}N_4O_4$ $[M]^+$: 334.1641, found: 334.1643.

^{1}H NMR (400 MHz, MeOD):

δ = 8.63 (s, 1H, *CH hAr*), 7.93 (d, J = 8.4 Hz, 2H, *CH Ar*), 7.56 (d, J = 8.4 Hz, 2H, *CH Ar*), 5.60 (d, J_{1-2} = 9.2 Hz, 1H, *H-1*), 4.22 (bs, 2H, *CH_2*), 4.15 (t, J_{2-1} = J_{2-3} = 9.3 Hz, 1H, *H-2*), 4.00 (qd, J_{5-4} = 1.1 Hz, J_{5-CH3} = 6.5 Hz, 1H, *H-5*), 3.80 (dd, J_{4-3} = 3.4 Hz, J_{4-5} = 1.1 Hz, 1H, *H-4*), 3.74 (dd, J_{3-2} = 9.5 Hz, J_{3-4} = 3.3 Hz, 1H, *H-3*), 2.75 (bs, 3H, *N-CH_3*), 1.32 (d, J_{CH3-5} = 6.5 Hz, 3H, *CH_3*).

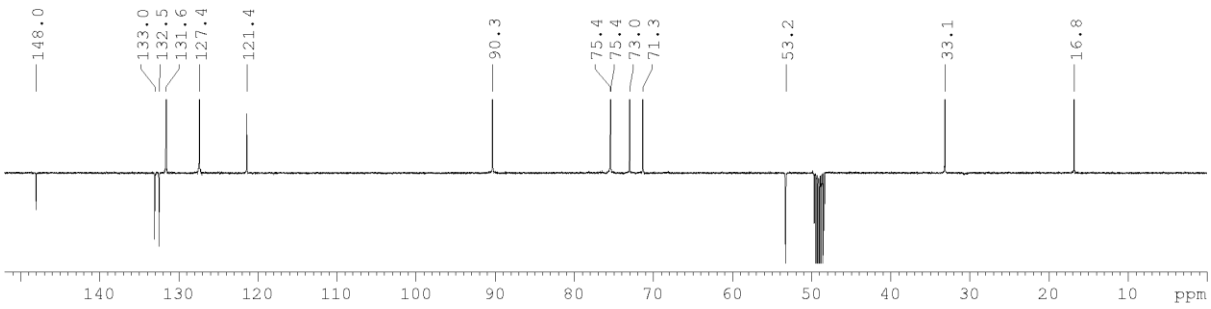


COSY:

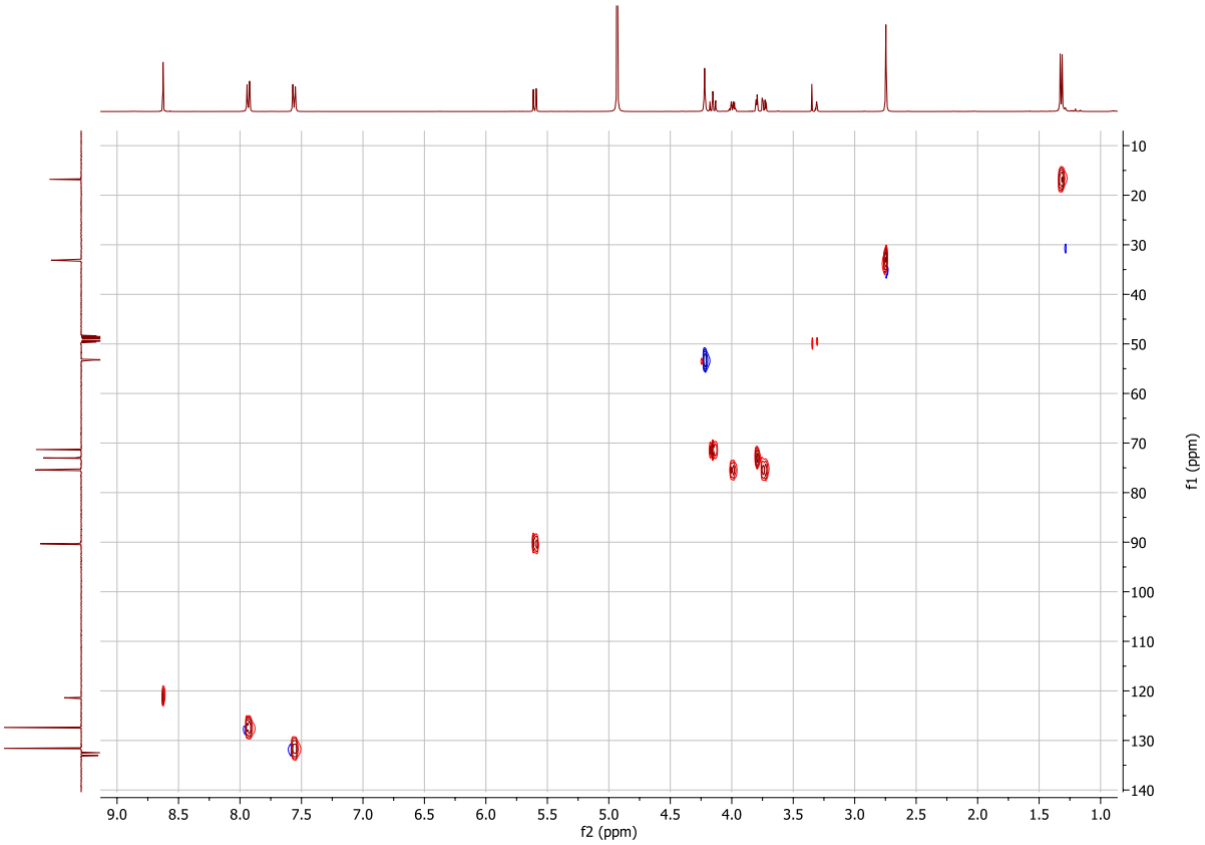


^{13}C NMR (400 MHz, MeOD):

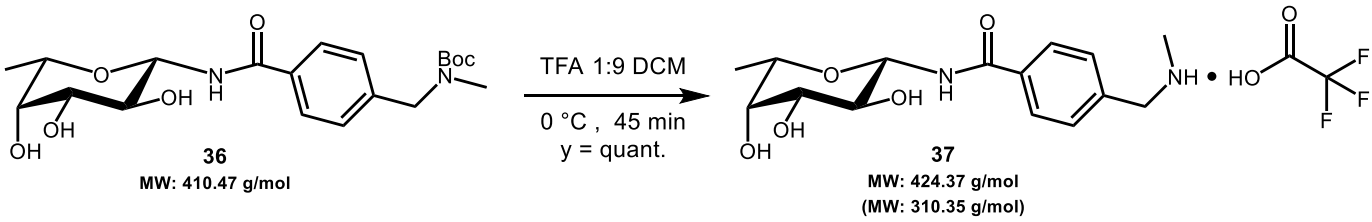
δ = 148.0 (*C hAr*), 133.0, 132.5 (C Ar), 131.6, 127.4 (*CH Ar*), 121.4 (*CH hAr*), 90.3 (*H1*), 75.4, 75.4 (*C3, C5*), 73.0 (*C4*), 71.3 (*C2*), 53.2 (CH_2), 33.1 (CH_3), 16.8 (*C6*).



HSQC:



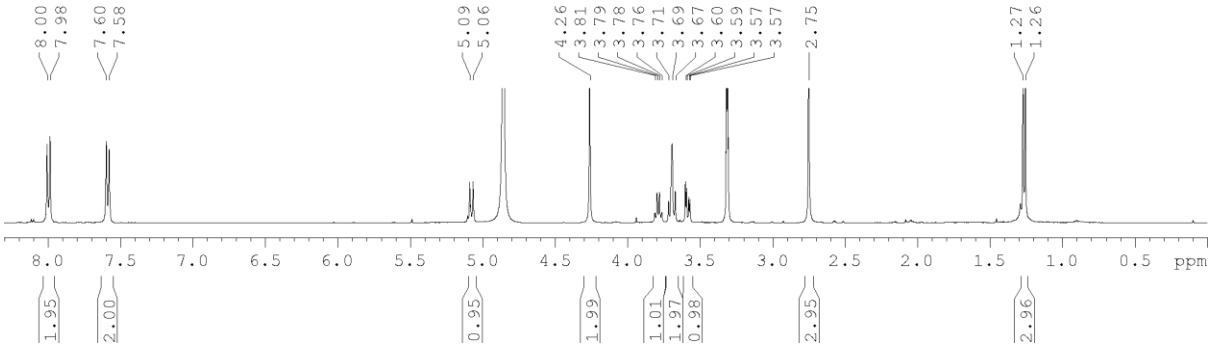
(37) Synthesis and characterization of **N-(β-L-fucopyranosyl)-4-((methylamino)methyl) benzamide (37)**:



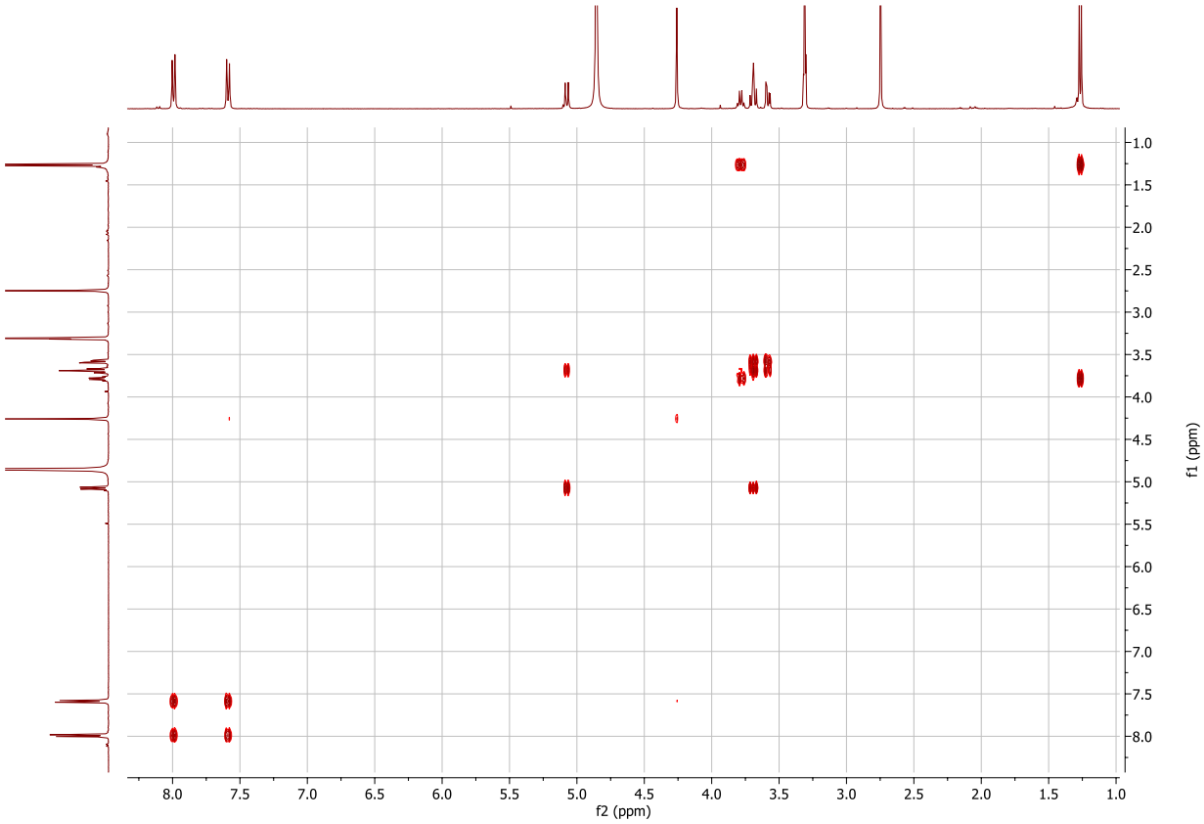
***tert*-butyl (4-((β-L-fucopyranosyl)carbamoyl)benzyl) (methyl) carbamate 36** (0.054 mmol) was subjected to the aforementioned procedure to afford **37** as a TFA salt (quantitative yield). TLC R_f (DCM/MeOH: 85/15): 0.13. $[\alpha]_D^{17}$ = -5.1 (MeOH, c 1). Exact Mass calculated for $C_{15}H_{22}N_2O_5$ $[M]^+$: 310.1529, found: 310.1514.

^{1}H NMR (400 MHz, MeOD):

δ = 7.99 (d, J = 8.5 Hz, 2H, *CH Ar*), 7.59 (d, J = 8.5 Hz, 2H, *CH Ar*), 5.08 (d, $J_{1\text{-}2}$ = 9.1 Hz, 1H, *H-1*), 4.26 (bs, 2H, *CH_2*), 3.79 (qd, $J_{5\text{-}4}$ = 1.1 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.69 (mult., $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.3 Hz, $J_{4\text{-}3}$ = 3.4 Hz, $J_{4\text{-}5}$ = 1.1 Hz, 2H, *H-2* + *H-4*), 3.58 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.4 Hz, 1H, *H-3*), 2.75 (bd, 3H, *N-CH_3*), 1.27 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*).

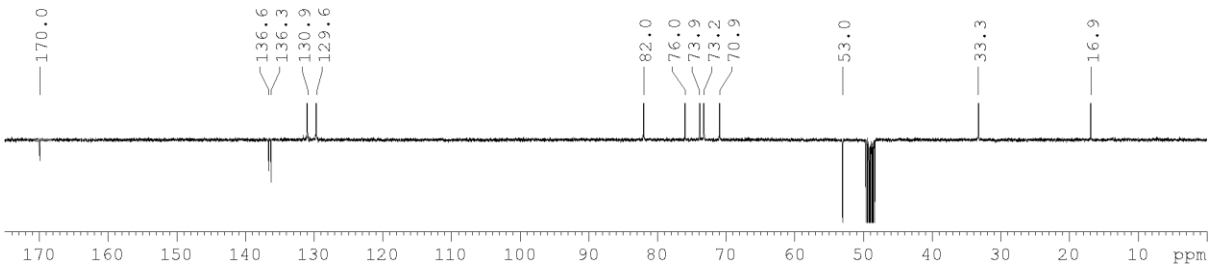


COSY:

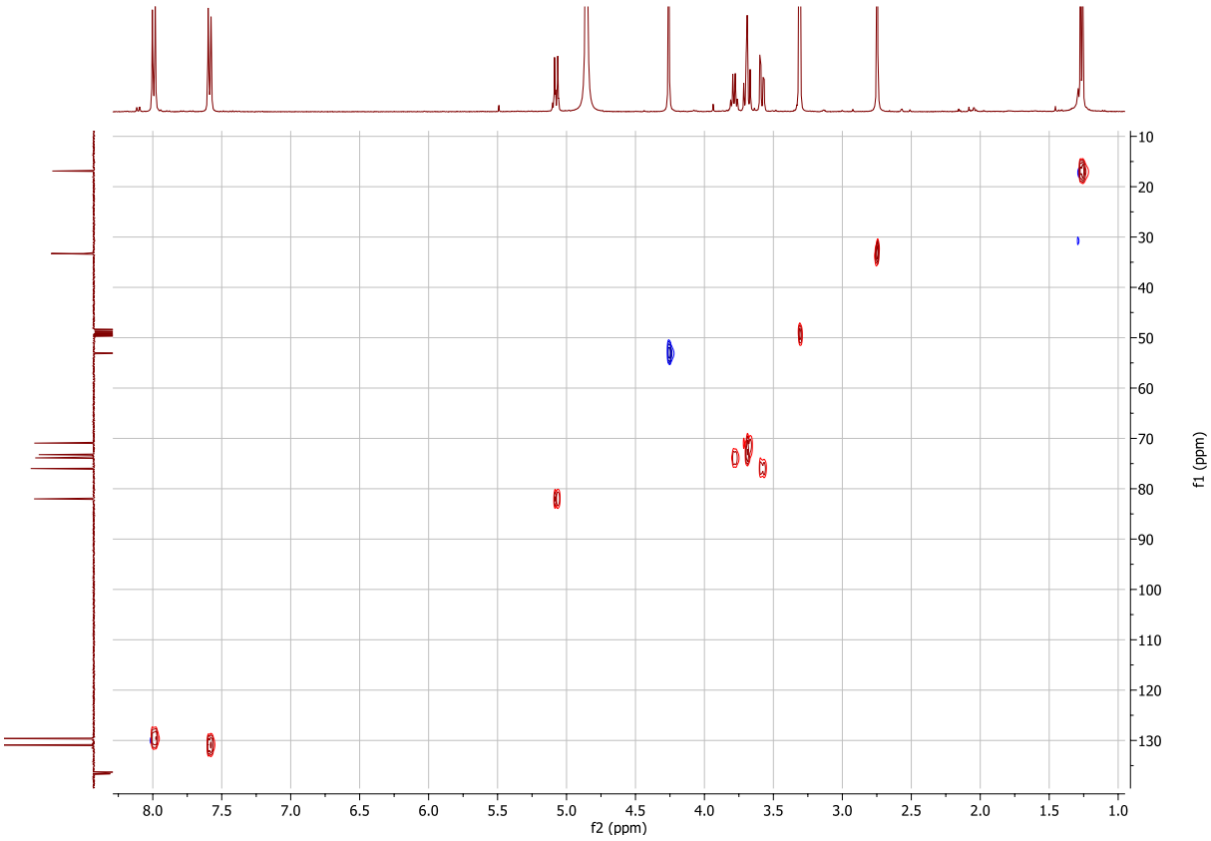


^{13}C NMR (400 MHz, MeOD):

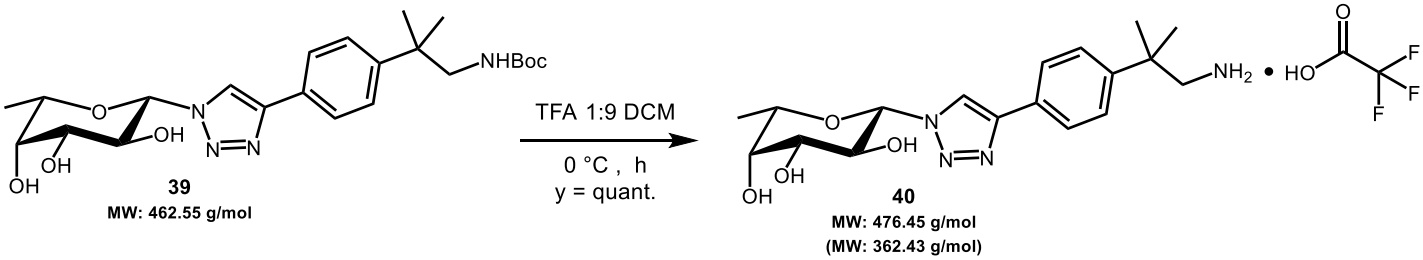
δ = 170.0 (*C=O*), 136.6, 133.3 (*C Ar*), 130.9, 129.6 (*CH Ar*), 82.0 (*H1*), 76.0 (*C3*), 73.9 (*C5*), 73.2 (*C4*), 70.9 (*C2*), 53.0 (CH_2), 34.3 (CH_3), 16.5 (*C6*).



HSQC:



(40) Synthesis and characterization of **2-(4-(1-(β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)phenyl)-2-methylpropan-1-amine (40)**:



***tert*-butyl (2-(4-(1-(β-L-fucopyranosyl)-1H-1,2,3-triazol-4-yl)phenyl)-2-methylpropyl) carbamate 39** (0.063 mmol) was subjected to the aforementioned procedure to afford **40** as a TFA salt (quantitative yield). TLC R_f (DCM/MeOH: 85/15): 0.25. Exact Mass calculated for $C_{18}H_{26}N_4O_4$ $[M]^+$: 362.1954, found 362.1948.

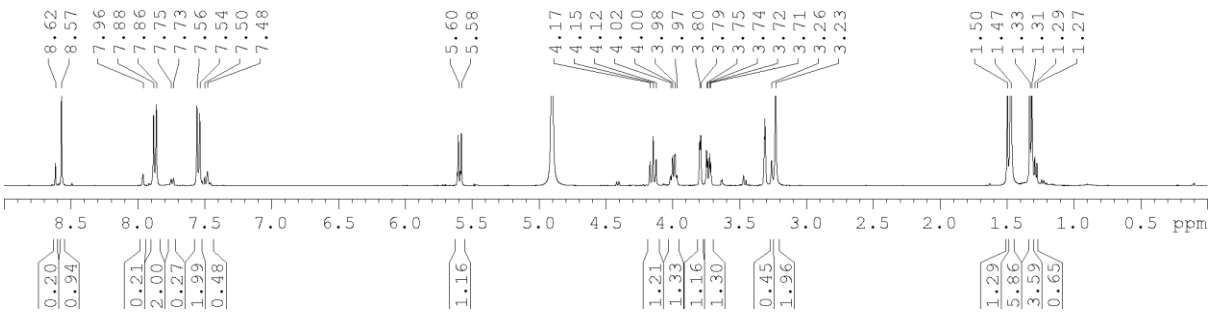
^{1}H NMR (400 MHz, MeOD): regioisomeric *para*/*meta* mixture (83/17)

para-isomer:

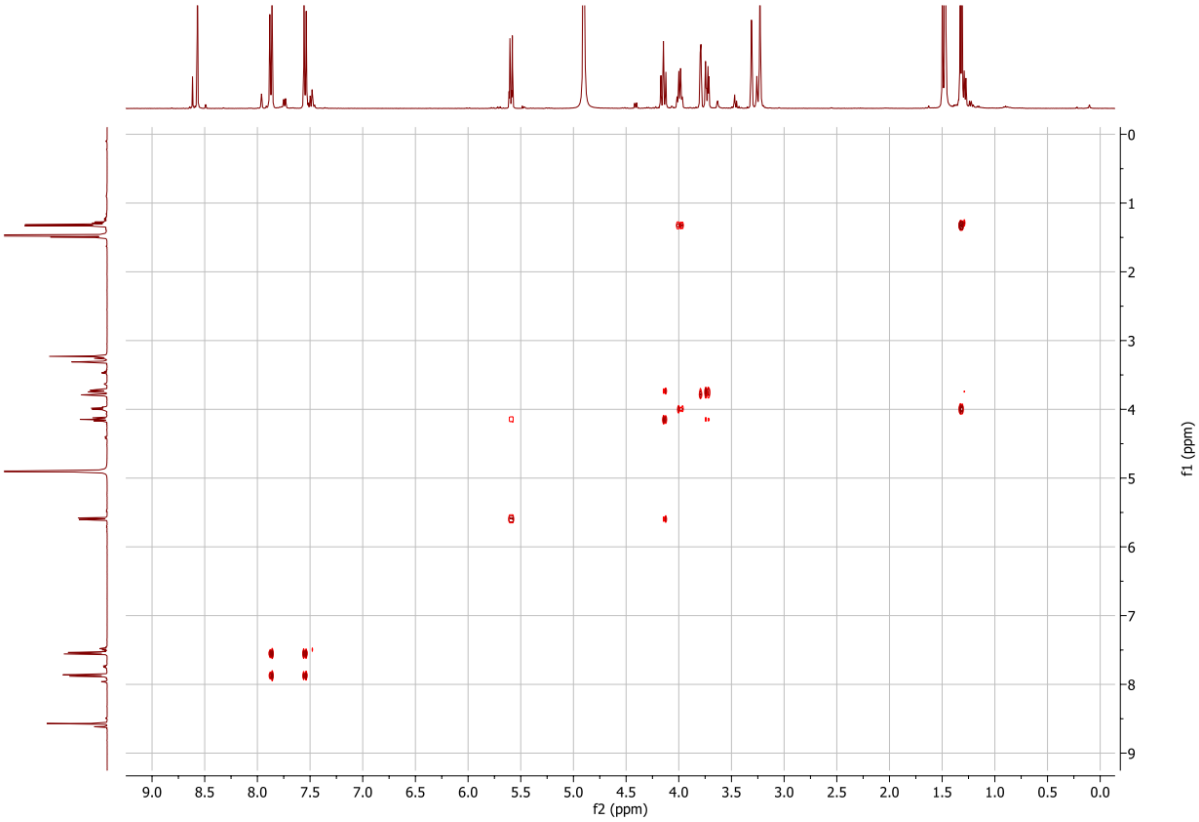
δ = 8.57 (s, 1H, *CH hAr*), 7.87 (d, J = 8.4 Hz, 2H, *CH Ar*), 7.55 (d, J = 8.5 Hz, 2H, *CH Ar*), 5.59 (d, $J_{1\text{-}2}$ = 9.2 Hz, 1H, *H-1*), 4.15 (t, $J_{2\text{-}1}$ = $J_{2\text{-}3}$ = 9.3 Hz, 1H, *H-2*), 3.99 (dq, $J_{5\text{-}4}$ = 1.0 Hz, $J_{5\text{-}6}$ = 6.5 Hz, 1H, *H-5*), 3.79 (d, $J_{4\text{-}3}$ = 3.2 Hz, $J_{4\text{-}5}$ = 1.0 Hz, 1H, *H-4*), 3.73 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.3 Hz, 1H, *H-3*), 3.23 (bs, 2H, CH_2), 1.47 (bs, 6H, *2x*CH_3), 1.32 (d, $J_{6\text{-}5}$ = 6.5 Hz, 6H, *H-6*).

meta-isomer:

δ = 8.62 (s, 1H, *CH hAr*), 7.96 (t, J = 1.7 Hz, 1H, *CH Ar*), 7.74 (dt, J' = 7.0 Hz, J = 1.7 Hz, 1H, *CH Ar*), 7.49 (m, 2H, *CH Ar*), 5.60 (d, $J_{1\text{-}2}$ = 9.2 Hz, 1H, *H-1*), 4.15 (t, 1H, *H-2*), 3.99 (dq, 1H, *H-5*), 3.79 (d, 1H, *H-4*), 3.73 (m, 1H, *H-3*), 3.26 (bs, 2H, CH_2), 1.50 (bs, 6H, *2x*CH_3), 1.28 (d, $J_{6\text{-}5}$ = 6.5 Hz, 6H, *H-6*).



COSY:



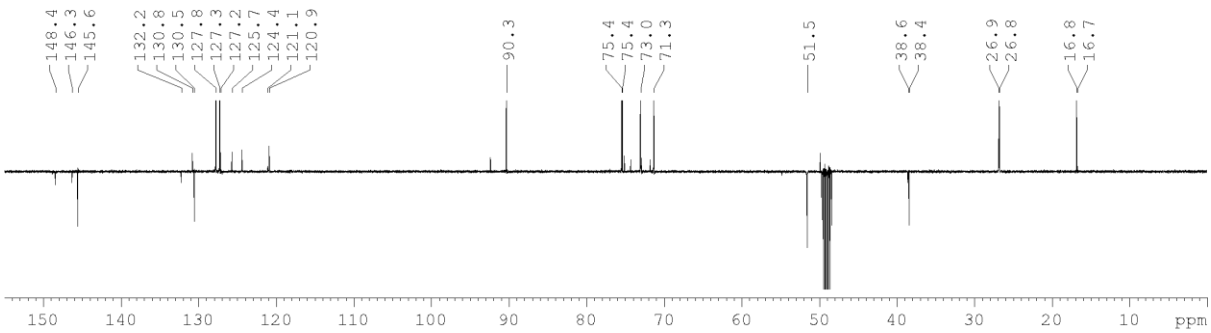
^{13}C NMR (400 MHz, MeOD):

para-isomer:

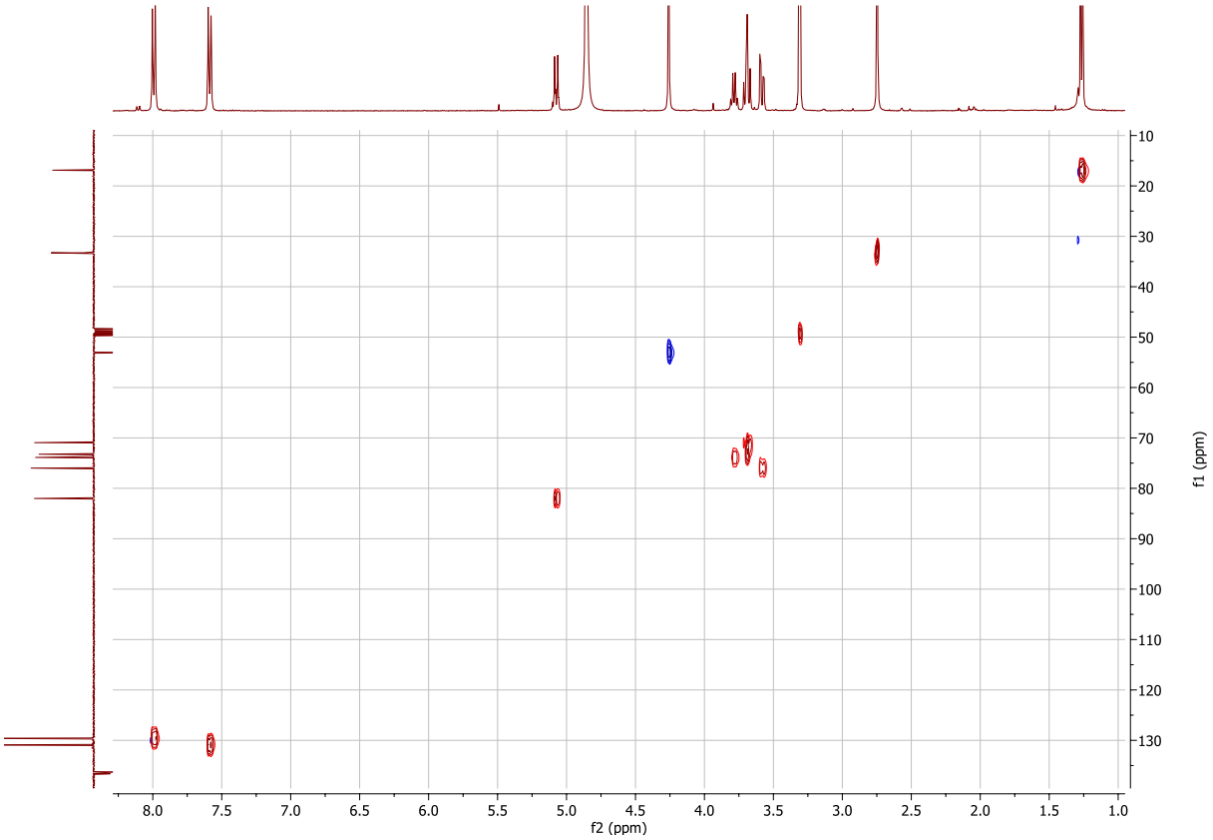
δ = 148.4 (*C hAr*), 145.6 (*C(CH$_3$)$_2$-C Ar*), 130.5 (*hAr-C Ar*), 127.8 (*CH Ar*), 127.3 (*CH Ar*), 120.9 (*CH hAr*), 90.3 (*C1*), 75.4, 75.4 (*C3* + *C5*), 73.0 (*C4*), 71.3 (*C2*), 51.5 (*CH$_2$*), 38.4 (*C(CH$_3$)$_2$*), 26.8 (2x*CH$_3$*), 16.8 (*C6*).

meta-isomer:

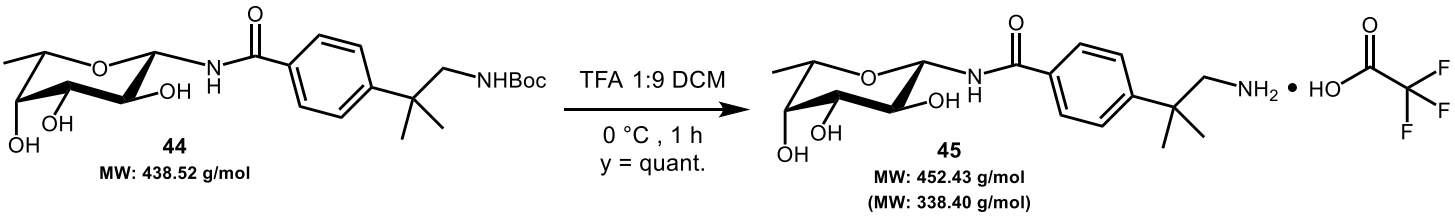
δ = 148.4 (*C hAr*), 146.3 (*C(CH$_3$)$_2$-C Ar*), 132.2 (*hAr-C Ar*), 130.7, 127.2, 125.7, 124.4 (*CH Ar*), 121.1 (*CH hAr*), 90.3 (*C1*), 75.4, 75.4 (*C3* + *C5*), 73.0 (*C4*), 71.3 (*C2*), 51.5 (*CH$_2$*), 38.6 (*C(CH$_3$)$_2$*), 26.9 (2x*CH$_3$*), 16.7 (*C6*).



HSQC:



(45) Synthesis and characterization of **4-(1-amino-2-methylpropan-2-yl)-N-(β-L-fucopyranosyl) benzamide (45)**:



***tert*-butyl (2-(4-((β-L-fucopyranosyl)carbamoyl)phenyl)-2-methylpropyl) carbamate 44** (0.059 mmol) was subjected to the aforementioned procedure to afford **45** as a TFA salt (quantitative yield). TLC R_f (DCM/MeOH: 85/15): 0.12. MS (ESI) calculated for Exact Mass calculated for $C_{17}H_{26}N_2O_5$ $[M]^+$: 338.1842, found 338.1827.

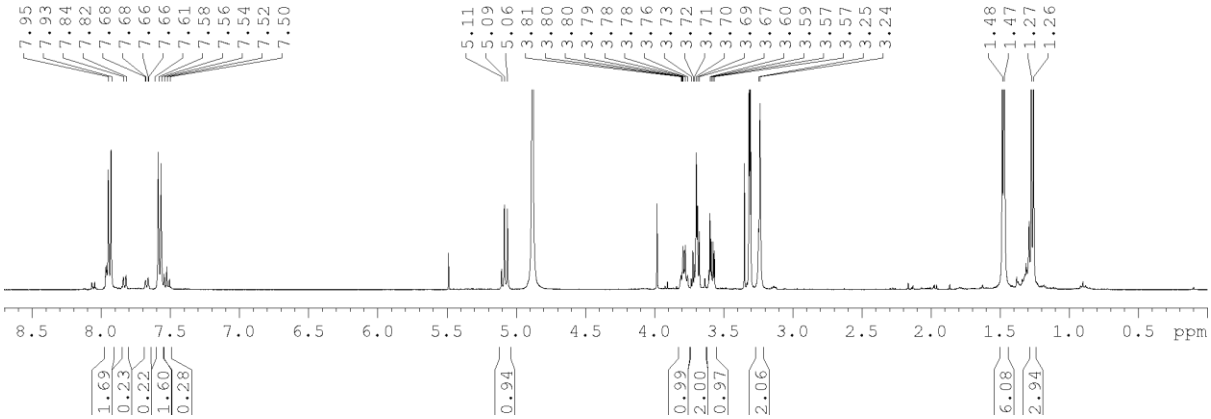
^{1}H NMR (400 MHz, MeOD): regioisomeric *para*/*meta* mixture (80/20)

para-isomer:

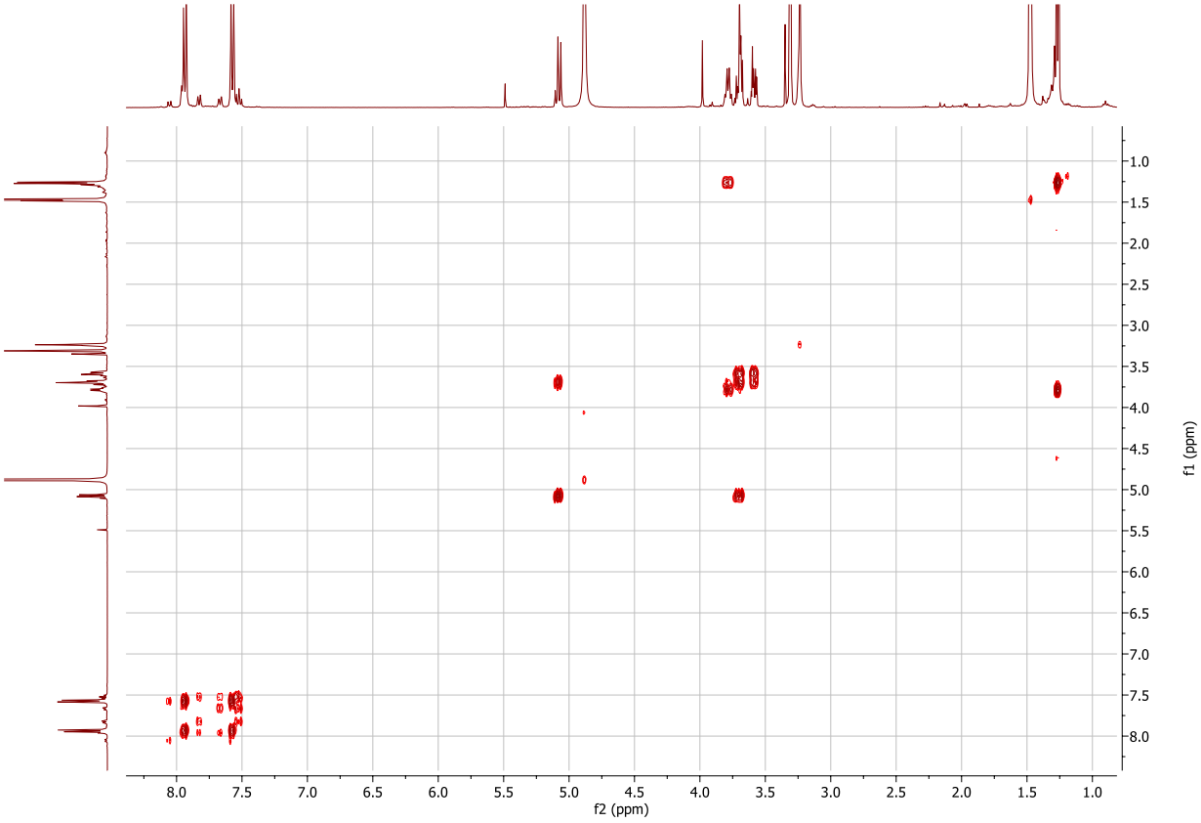
δ = 7.94 (d, J = 8.7 Hz, 2H, *CH Ar*), 7.57 (d, J = 8.7 Hz, 2H, *CH Ar*), 5.07 (d, $J_{1\text{-}2}$ = 9.1 Hz, 1H, *H-1*), 3.78 (dq, $J_{5\text{-}4}$ = 1.1 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.70 (mult., $J_{2\text{-}1}$ = 9.1 Hz, $J_{2\text{-}3}$ = 9.4 Hz, $J_{4\text{-}3}$ = 3.3 Hz, $J_{4\text{-}5}$ = 1.1 Hz, 2H, *H-2* + *H-4*), 3.58 (dd, $J_{3\text{-}2}$ = 9.5 Hz, $J_{3\text{-}4}$ = 3.3 Hz, 1H, *H-3*), 3.24 (bs, 2H, CH_2), 1.47 (bs, 6H, 2xCH_3), 1.27 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *H-6*).

meta-isomer:

δ = 7.95 (m, 1H, *CH Ar*), 7.83 (ddd, J = 7.7 Hz, J = 1.7 Hz, J = 1.1 Hz, 1H, *CH Ar*), 7.67 (ddd, J = 7.9 Hz, J = 2.1 Hz, J = 1.1 Hz, 1H, *CH Ar*), 7.52 (t, J = 7.8 Hz H, *CH Ar*), 5.09 (d, $J_{1\text{-}2}$ = 9.1 Hz, 1H, *H-1*), 3.79 (dq, 1H, *H-5*), 3.70 (mult., 2H, *H-2* + *H-4*), 3.58 (dd, 1H, *H-3*), 3.25 (bs, 2H, CH_2), 1.48 (bs, 6H, 2xCH_3), 1.27 (d, 3H, *H-6*).



COSY:



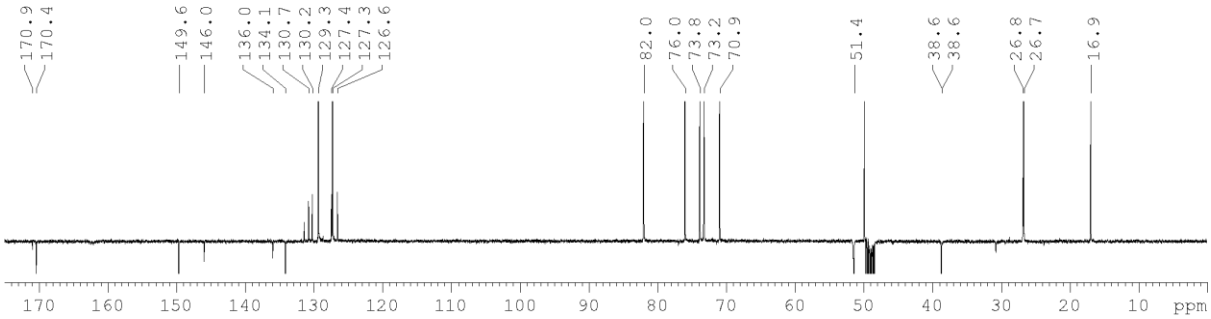
^{13}C NMR (400 MHz, MeOD):

para-isomer:

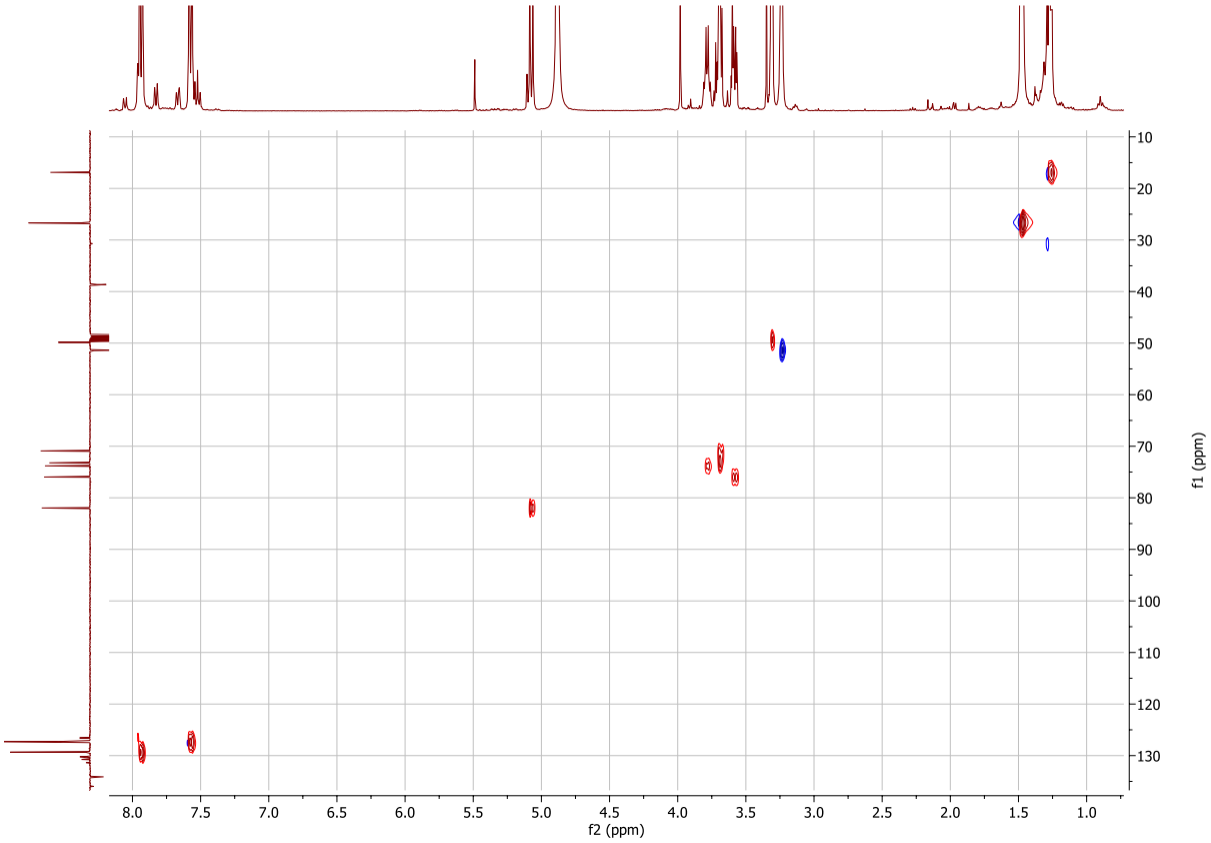
δ = 170.4 (*C=O*), 149.6 (*C(CH₃)₂-C Ar*), 134.1 (*C Ar*), 129.3 (*CH Ar*), 127.3 (*CH Ar*), 82.0 (*C1*), 76.0 (*C3*), 73.8 (*C5*), 73.2 (*C4*), 70.9 (*C2*), 51.4 (CH_2), 38.6 (*C(CH₃)₂*), 26.7 (2xCH_3), 16.9 (*C6*).

meta-isomer:

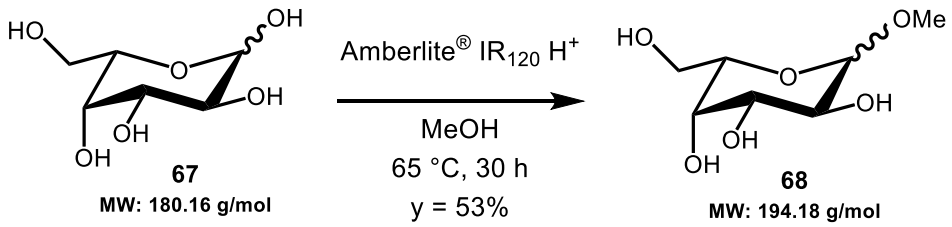
δ = 170.9 (*C=O*), 146.0 (*C(CH₃)₂-C Ar*), 136.0 (*C Ar*), 130.7, 130.2, 127.4, 126.6 (*C Ar*), 82.0 (*C1*), 76.0 (*C3*), 73.8 (*C5*), 73.2 (*C4*), 70.9 (*C2*), 51.4 (CH_2), 38.6 (*C(CH₃)₂*), 26.8 (2xCH_3), 16.9 (*C6*).



HSQC:



(68) Synthesis and characterization of **Methyl α/β-L-galactopyranoside (68)** following the procedure of Wang and co-workers:[213, 235]

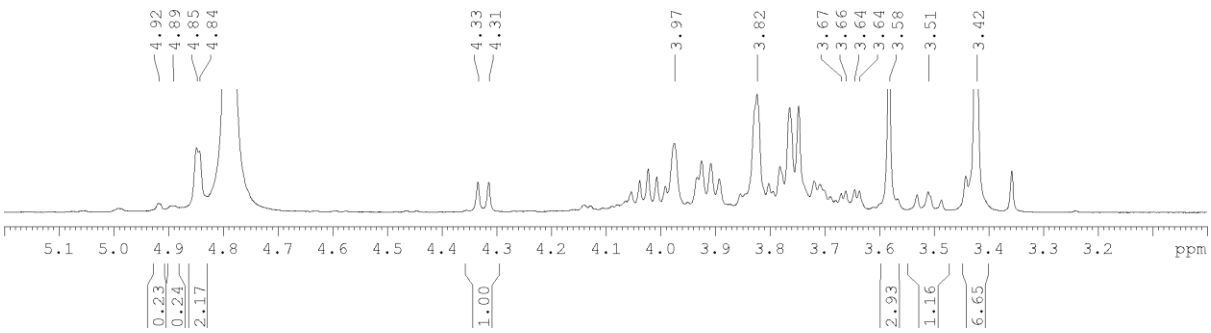


To a solution of L-galactose **67** (98 mg, 0.54 mmol, 1 eq) in dry MeOH (1.5 mL) was added Amberlite® IR_{120} H^+ resin (102 mg) under N_2 atmosphere. The reaction mixture was heated to reflux (65 °C) and stirred for 30 h until TLC showed no further advancement, before returning to room temperature and filtering. Evaporation afforded a mixture of Me-α-L-galactopyranoside, Me-β-L-galactopyranoside, Me-α-L-galactofuranoside, and Me-β-L-galactofuranoside in ratio 56/19/8/17. Recrystallization of Me-α/β-L-galactopyranoside **68** was achieved by redissolving the crude in iPrOH, refluxing (83 °C) and cooling to -17 °C, then filtering white crystals (6 mg, 0.03 mmol, y = 6 %). The remaining mother liquor was recycled by re-equilibrating in refluxing MeOH (65 °C) with Amberlite® IR_{120} H^+ resin for 48 h. Further recrystallization and recycling increased the final yield to y = 56 %. TLC R_f (DCM/MeOH: 7/3): 0.38 (Me-α-L-galactopyranoside), 0.48 (Me-β-L-galactopyranoside).

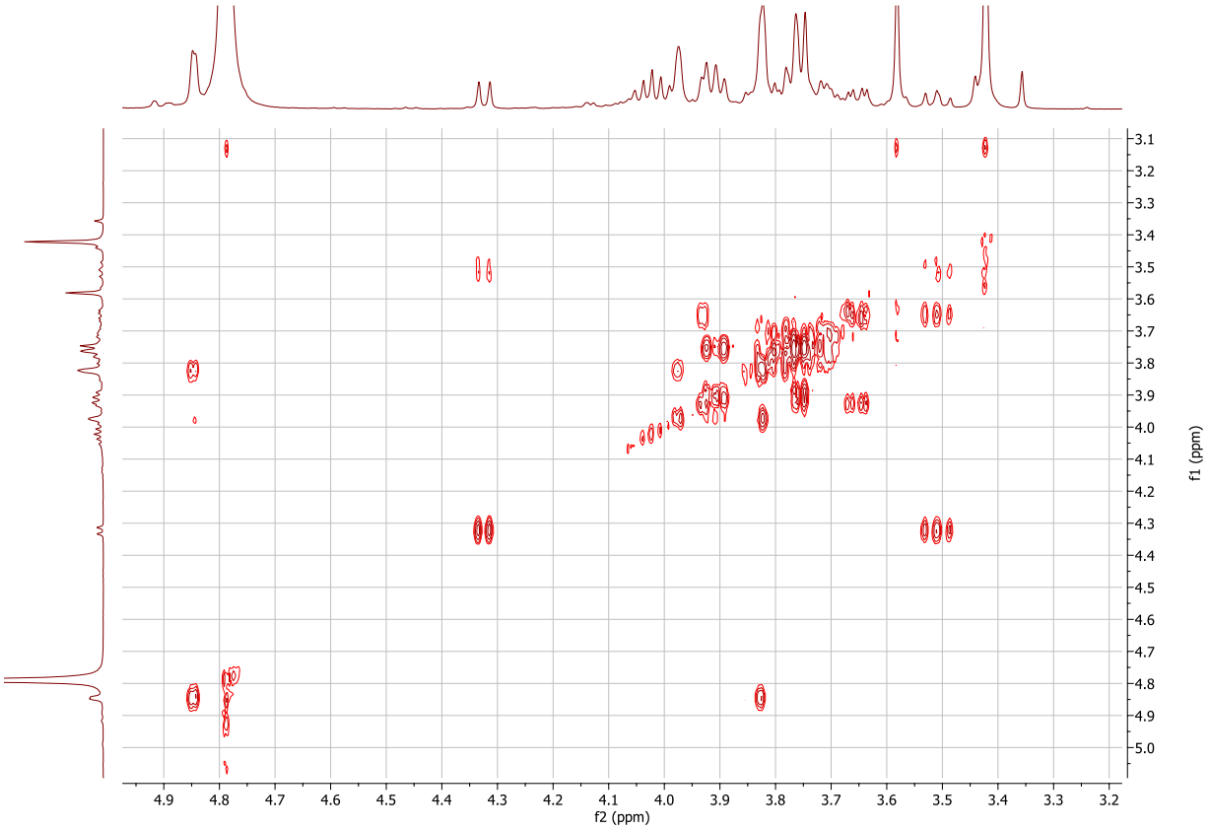
^{1}H NMR (400 MHz, D_2O):

Crystals of Me-α/β-L-fucopyranoside **68** (isomer ratio: 61/28/6/6). Some shifts were extrapolated from the COSY experiment:

δ = 4.85 (d, α*H-1*), 4.32 (d, β*H-1*), 3.97 (m, α*H-4*), 3.82 (m, α*H-2*), 3.58 (s, βOCH_3), 3.51 (t, β*H-2*), 3.43 (s, αOCH_3).

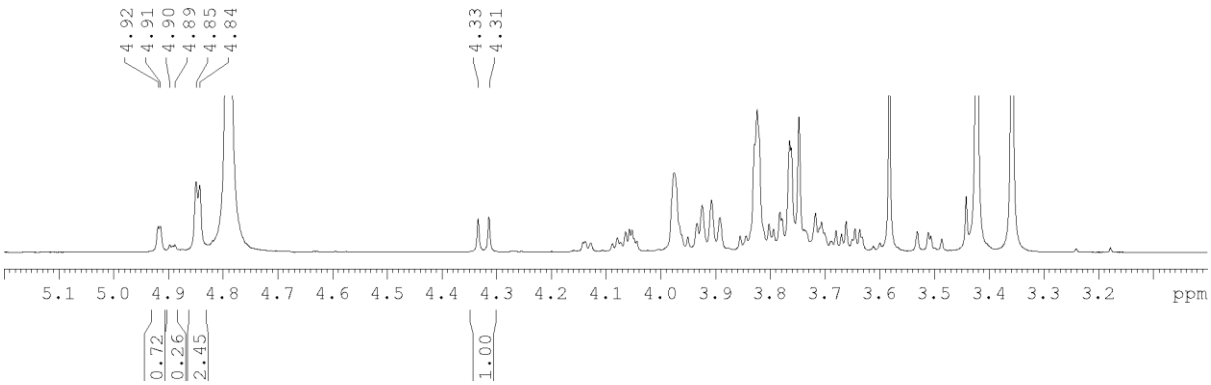


COSY:

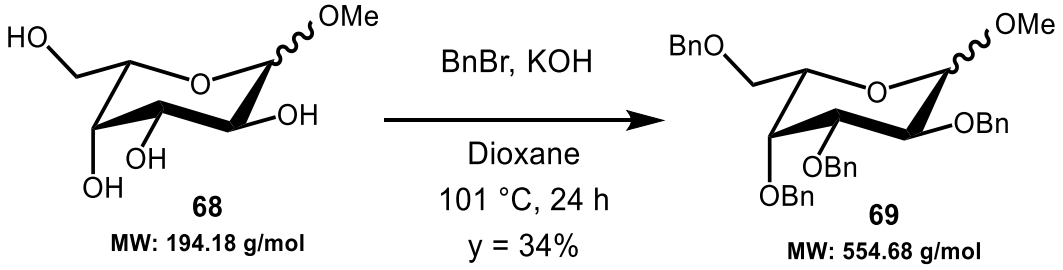


Crude mixture of 4 isomers:

δ = 4.92 (**β furanoside**, d, $J_{1\text{-}2}$ = 1.7 Hz, *H-1*), 4.89 (**α furanoside**, *H-1*), 4.84 (**α pyranoside**, d, $J_{1\text{-}2}$ = 2.7 Hz, *H-1*), 4.32 (**β pyranoside**, d, $J_{1\text{-}2}$ = 7.9 Hz, *H-1*). Calculated ratio/percentage: 56/19/8/17, in accordance with published data.[235]



(69) Synthesis and characterization of **Methyl 2,3,4,6-tetra-O-benzyl L-galactopyranoside (69):**

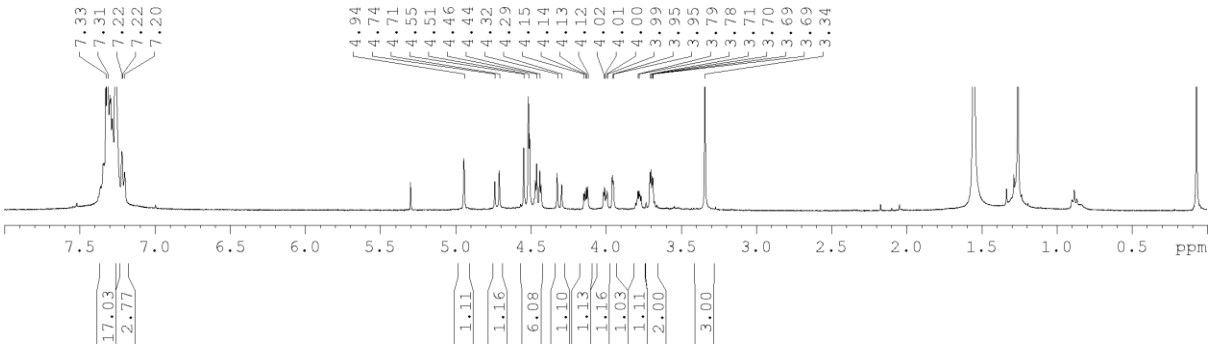


To a solution of **methyl α/β-L-galactopyranoside 68** (600 mg, 3.09 mmol, 1 eq) in dry Toluene (30 mL) was added BnBr (5.9 mL, 49.61 mmol, 16 eq), followed by ground KOH (1.8 g, 32.08 mmol, 10 eq). The reaction mixture was heated to reflux (101 °C) under N_2 atmosphere and stirred for 24 h until TLC showed no further advancement, before returning to room temperature. The mixture was quenched with ice-cold water and extracted with EtOAc. The organic phase was washed with ice-cold water and dried over Na_2SO_4. The crude was purified by automatic chromatography (Biotage SNAP 100: nHex/EtOAc 5 % isocratic, then gradient to 80 %) affording product **69** (583 mg, 1.05 mmol, y = 34 %). TLC R_f (nHex/EtOAc: 8/2): 0.31 (α-L-galactopyranoside), 0.36 (β-L-galactopyranoside).

α-anomer:

^{1}H NMR (400 MHz, $CDCl_3$):

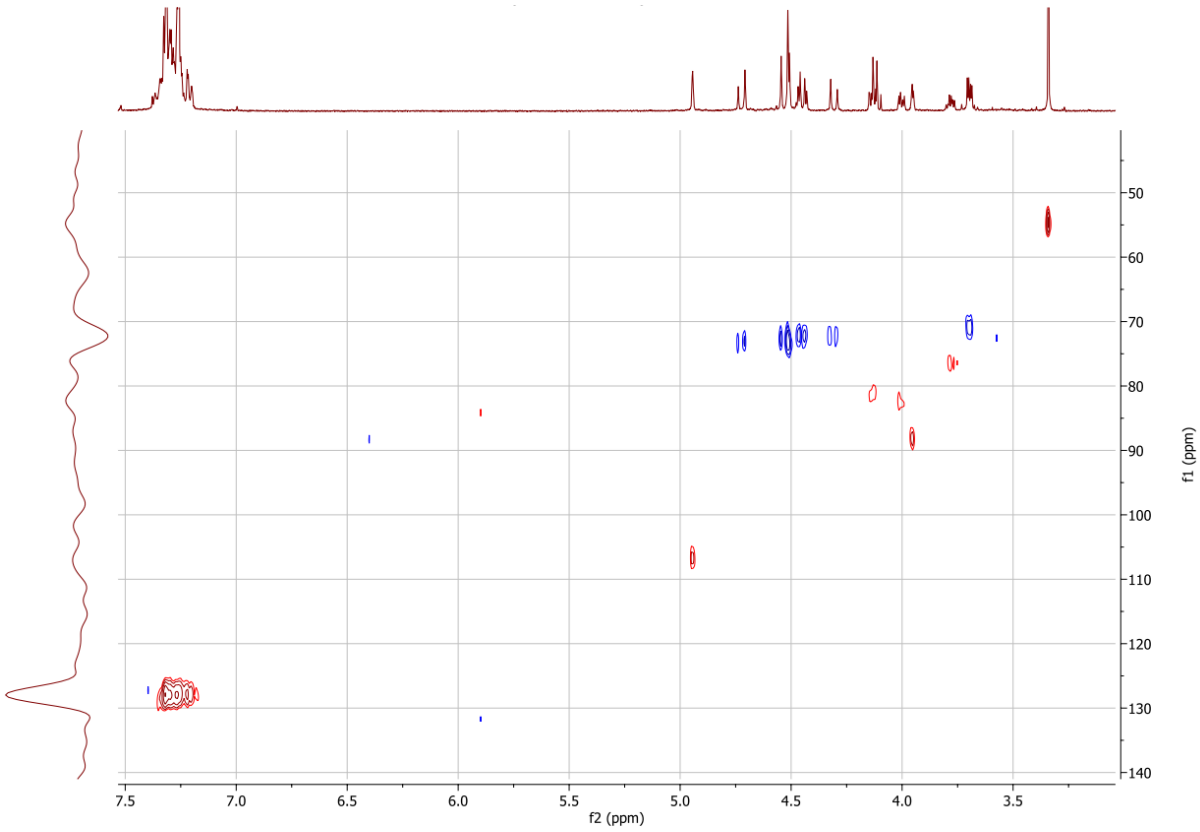
δ = 7.33 - 7.20 (m, 20H, *Ar*), 4.94 (bs, 1H, *H-1*), 4.72 (d, 1H, *CH_2-OBn*), 4.55 - 4.44 (mult., 6H, *CH_2-OBn*), 4.31 (d, 1H, *CH_2-OBn*), 4.13 (dd, J_{2-1} = 3.4 Hz, J_{2-3} = 6.9 Hz, 1H, *H-2*), 4.00 (dd, J_{3-2} = 6.9 Hz, J_{3-4} = 3.0 Hz, 1H, *H-3*), 3.95 (dd, J_{4-3} = 3.0 Hz, J_{4-5} = 1.3 Hz, 1H, *H-4*), 3.78 (m, J_{5-6} = 5.9 Hz, 1H, *H-5*), 3.70 (dd, J_{6-5} = 5.9 Hz, 2H, *H-6*), 3.34 (s, 3H, *OCH_3*).



^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 128.0 (*CH Ar*), 106.9 (*C-1*), 88.2 (*C-4*), 82.7 (*C-3*), 81.2 (*C-2*), 76.5 (*C-5*), 73.2 – 72.0 (*CH_2-OBn*), 66.1 (*C-5*), 71.0 (C-6), 54.7 (*OCH_3*).

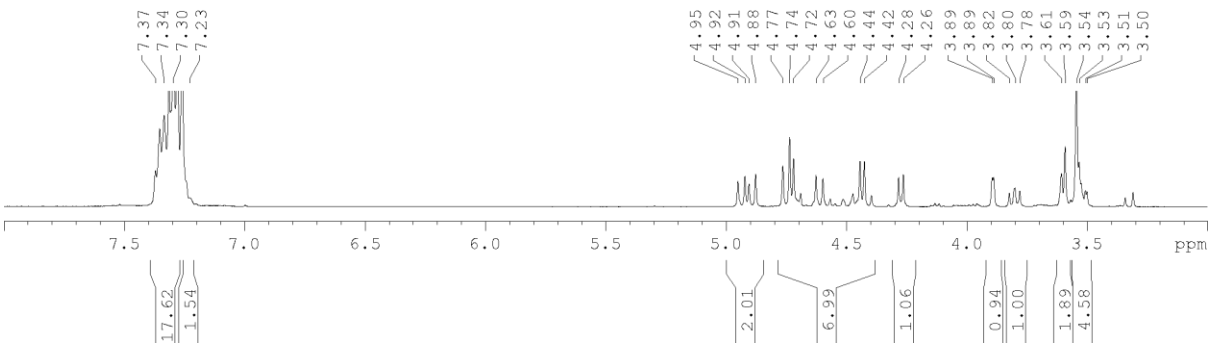
HSQC:



β-anomer:

^{1}H NMR (400 MHz, $CDCl_3$):

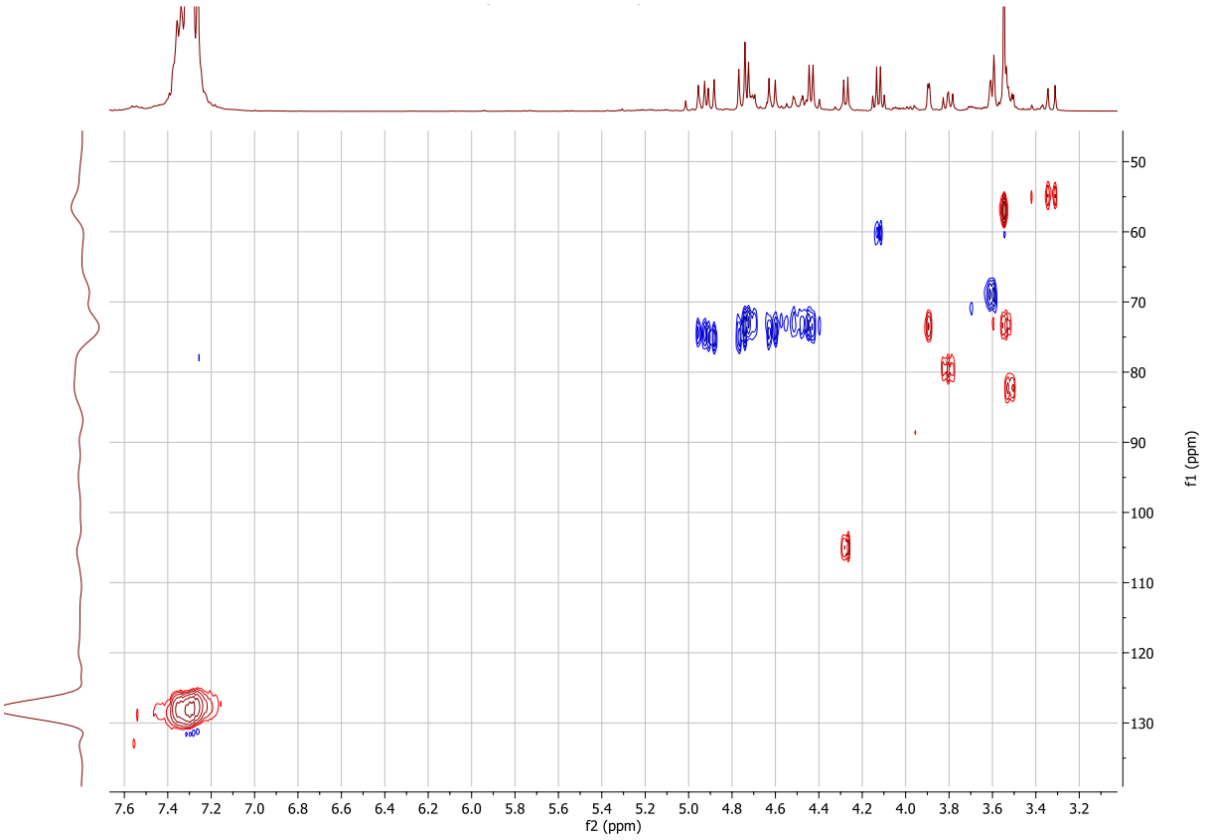
δ = 7.37 - 7.23 (m, 20H, *Ar*), 4.95 - 4.42 (mult., 8H, *CH_2-OBn*), 4.27 (d, J_{1-2} = 7.7 Hz, 1H, *H-1*), 3.89 (d, J_{4-3} = 2.9 Hz, 1H, *H-4*), 3.80 (dd, J_{2-1} = 7.8 Hz, J_{2-3} = 9.8 Hz, 1H, *H-2*), 3.60 (m, 2H, *H6*), 3.54 (mult., 5H, *OCH_3* + *H-3* + *H-5*).



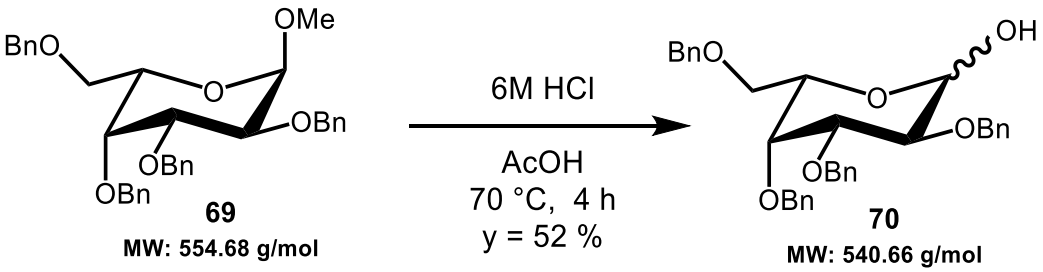
^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 128.1 (*CH Ar*), 105.0 (*C-1*), 82.2 (*C-3*), 79.7 (*C-2*), 75.2 – 72.9 (*CH_2-OBn*), 73.5 (*C-4*), 73.4 (*C-5*), 68.9 (C-6), 57.0 (*OCH_3*).

HSQC:



(70) Synthesis and characterization of **2,3,4,6-tetra-O-benzyl L-galactopyranoside (70)** following the procedure of Shi and co-workers:[236]



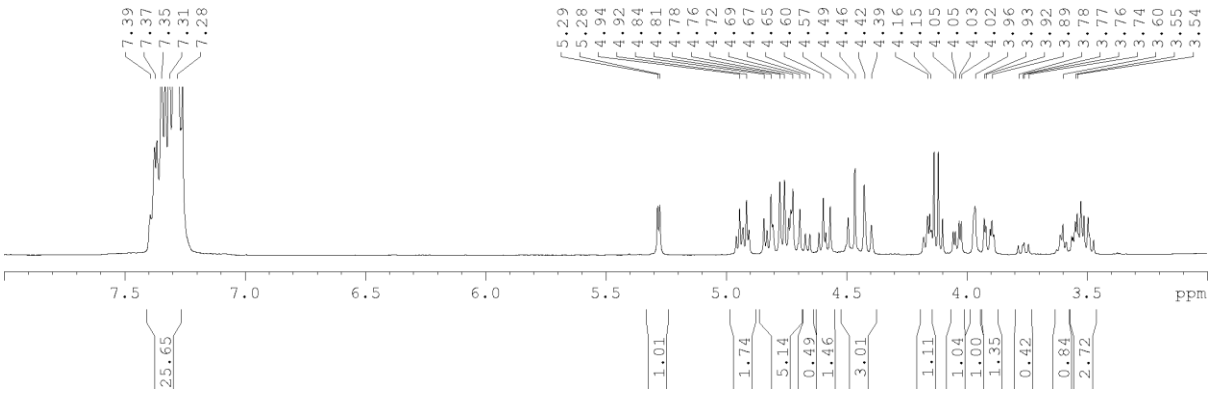
To a solution of **methyl 2,3,4,6-tetra-O-benzyl L-galactopyranoside 69** (590 mg, 1.06 mmol, 1 eq) in Acetic acid (3.6 mL) was added 6M HCl (585 µL, 3.51 mmol, 3.3 eq). The reaction mixture was heated to 70 °C and stirred for 4 h until TLC showed completion, before returning to room temperature. The mixture was quenched with ice-cold water and extracted with DCM. The organic phase was washed with $NaHCO_3$ aqueous solution and brine, then dried over $MgSO_4$. Recrystallization was achieved by redissolving the crude in a nHex/Et_2O solution

(4/2.4 mL: *ca.* 90 mg/mL), heating to 45 °C and cooling to -16 °C, then filtering white crystals of the anomeric mixture **70** (300 g, 0.56 mmol, y = 52 %). TLC R_f (nHex/EtOAc: 7/3): 0.20.

^{1}H NMR (400 MHz, $CDCl_3$): anomeric mixture α/β (ratio 3:1)

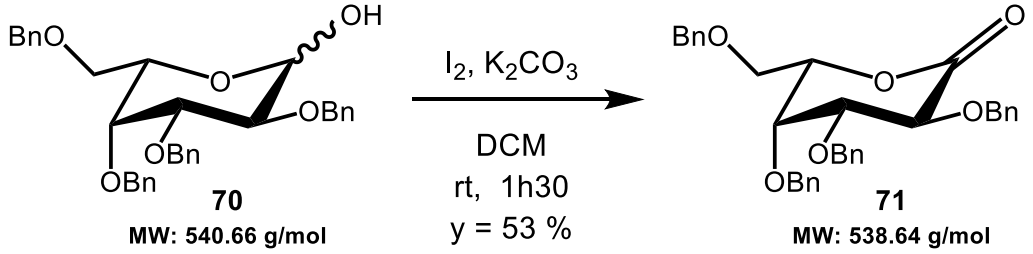
α-anomer: δ = 7.39 - 7.28 (m, 20H, *Ar*), 5.28 (d, J_{1-2}= 3.6 Hz, 1H, *H-1*), 4.96-4.39 (mult., 8H, *CH_2-OBn*), 4.16 (m, J_{5-6} = 6.4 Hz, 1H, *H-5*), 4.04 (dd, J_{2-1}=3.6 Hz, J_{2-3}=9.8 Hz, 1H, *H-2*), 3.96 (bs, 1H, *H-4*), 3.91 (m, J_{3-2} = 9.8 Hz, J_{3-4} = 2.6 Hz, 1H, *H-3*), 3.55-3.47 (m, 2H, *CH_2*).

β-anomer: δ = 7.39 - 7.28 (m, 20H, *Ar*), 4.66 (d, J_{1-2}= 7.4 Hz, 1H, *H-1*), 4.96-4.39 (mult., 8H, *CH_2-OBn*), 3.89 (m, 1H), 3.76 (dd, J_{2-1} = 7.6 Hz, J_{2-3} = 9.6 Hz, 1H, *H-2*), 3.60 (m, 2H, *H-6*), 3.55-3.47 (m, 2H, *H-6'*).



HSQC:

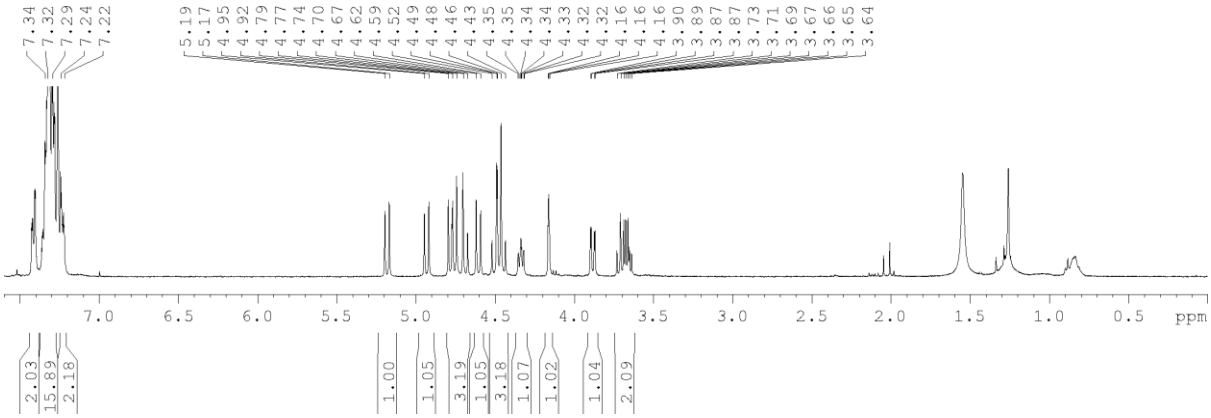
(71) Synthesis and characterization of **2,3,4,6-tetra-O-benzyl L-galactonolactone (71)** following the procedure of Fusaro and co-workers:[219]



To a solution of **2,3,4,6-tetra-O-benzyl L-galactopyranoside 70** (122 mg, 0.23 mmol, 1 eq) in DCM (1.4 mL) was added I_2 (205 mg, 0.81 mmol, 3.6 eq), followed by ground K_2CO_3 (110 mg, 0.80 mmol, 3.5 eq). The brown reaction mixture was stirred at room temperature for 1h30 before being quenched with ice-cold water and extracted with DCM. The organic phase was washed with $Na_2S_2O_3$ aqueous solution, becoming clear, and with brine and then dried over Na_2SO_4. The crude was purified by automatic chromatography (Biotage SNAP 100: nHex/EtOAc gradient from 2% to 30%) affording product **71** (64 mg, 0.12 mmol, y = 53 %). TLC R_f (nHex/EtOAc: 8/2): 0.34.

^{1}H NMR (400 MHz, $CDCl_3$):

δ = 7.41 (mult., 2H, *Ar*), 7.36 - 7.22 (mult., 18H, *Ar*), 5.18 (d, J= 11.0 Hz, 1H, *CH_2-OBn*), 4.93 (d, J'= 11.3 Hz, 1H, *CH_2-OBn*), 4.74 (mult., 3H, *CH_2-OBn*), 4.60 (d, J'= 11.3 Hz, 1H, *CH_2-OBn*), 4.47 (mult., 3H, $J_{2\text{-}3}$ = 9.5 Hz, *CH_2-OBn* + *H-2*), 4.34 (dq, $J_{5\text{-}4}$ = 1.6 Hz, $J_{5\text{-}6}$ = 5.7 Hz, 1H, *H-5*), 4.16 (dd, $J_{4\text{-}3}$ = $J_{4\text{-}5}$ = 1.9 Hz, 1H, *H-4*), 3.88 (dd, $J_{3\text{-}2}$ = 9.6 Hz, $J_{3\text{-}4}$ = 2.2 Hz, 1H, *H-3*), 3.68 (m, $J_{6\text{-}5}$ = 5.6 Hz, 2H, *CH_2*).

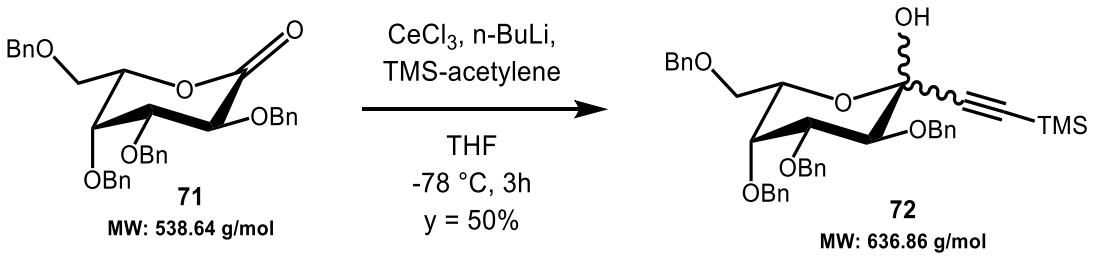


^{13}C chemical shifts were extrapolated from the HSQC experiment:

δ = 128.2 (*CH Ar*), 80.1 (*C-3*), 77.3 (*C-2*), 77.2 (*C-5*), 75.3, 74.8, 73.6, 72.8 (*CH_2-OBn*), 72.5 (*C-4*), 67.4 (C-6).

HSQC:

(72) Synthesis and characterization of **(1-hydroxy-2,3,4,6-tetra-O-benzyl L-galactopyranosyl) trimethylsilyl acetylene (72)** following the procedure of Lowary and co-workers:[181]



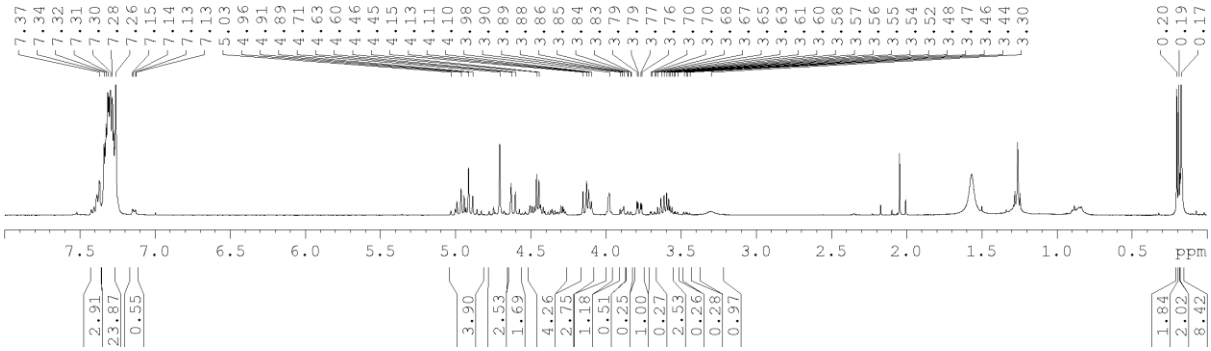
The procedure described for the synthesis of compound **6** was applied to **2,3,4,6-tetra-O-benzyl L-galactonolactone 71** (115 mg, 0.214 mmol, 1 eq) to afford **72** (68 mg, 0.107 mmol, y = 50 %). TLC R_f (nHex/tBuOMe: 7.5/2.5): 0.23.

^{1}H NMR (400 MHz, $CDCl_3$): anomeric ratio 1:0.27

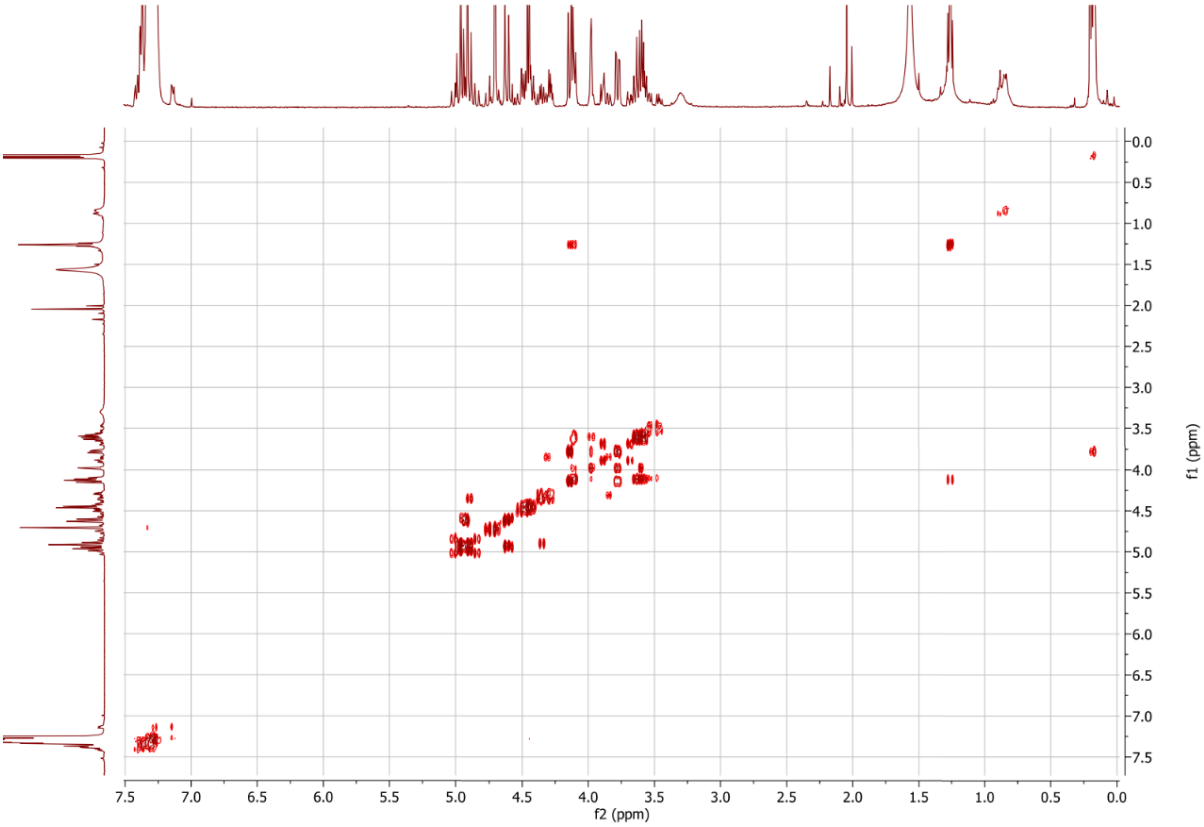
major-anomer: δ = 7.39 - 7.26 (m, 20H, *Ar*), 4.99 - 4.45 (mult., 8H, *CH_2-OBn*), 4.14 (d, J_{2-3}= 9.8 Hz, 1H, *H-2*), 4.11 (m, J_{5-4} = 1.2 Hz, J_{5-6} = 7.2 Hz, 1H, *H-5*), 3.98 (dd, J_{4-3} = 2.8 Hz, J_{4-5} = 1.2 Hz,

1H, *H-4*), 3.78 (dd, J_{3-2} = 9.7 Hz, J_{3-4} = 2.8 Hz, 1H, *H-3*), 3.65-3.56 (m, 2H, *CH_2*), 3.30 (bs, 1H, *OH*), 0.20 - 0.17 (m, 9H, *Si-CH_3*).

minor-anomer: δ = 7.42 - 7.26 (mult., 13H, *Ar*), 7.14 (dd, 2H, *Ar*), 5.03 - 4.27 (mult., 8H, *CH_2-OBn*), 3.89 (mult., J_{2-3} = 9.8 Hz, 2H, *H-2* + *H-4*), 3.84 (dd, J_{5-4} = 1.7 Hz, J_{5-6} = 8.3 Hz, 1H, *H-5*), 3.69 (dd, J_{3-2} = 10.2 Hz, J_{3-4} = 2.8 Hz, 1H, *H-3*), 3.55 -3.44 (m, 2H, *CH_2*), 0.20 - 0.17 (m, 9H, *Si-CH_3*).



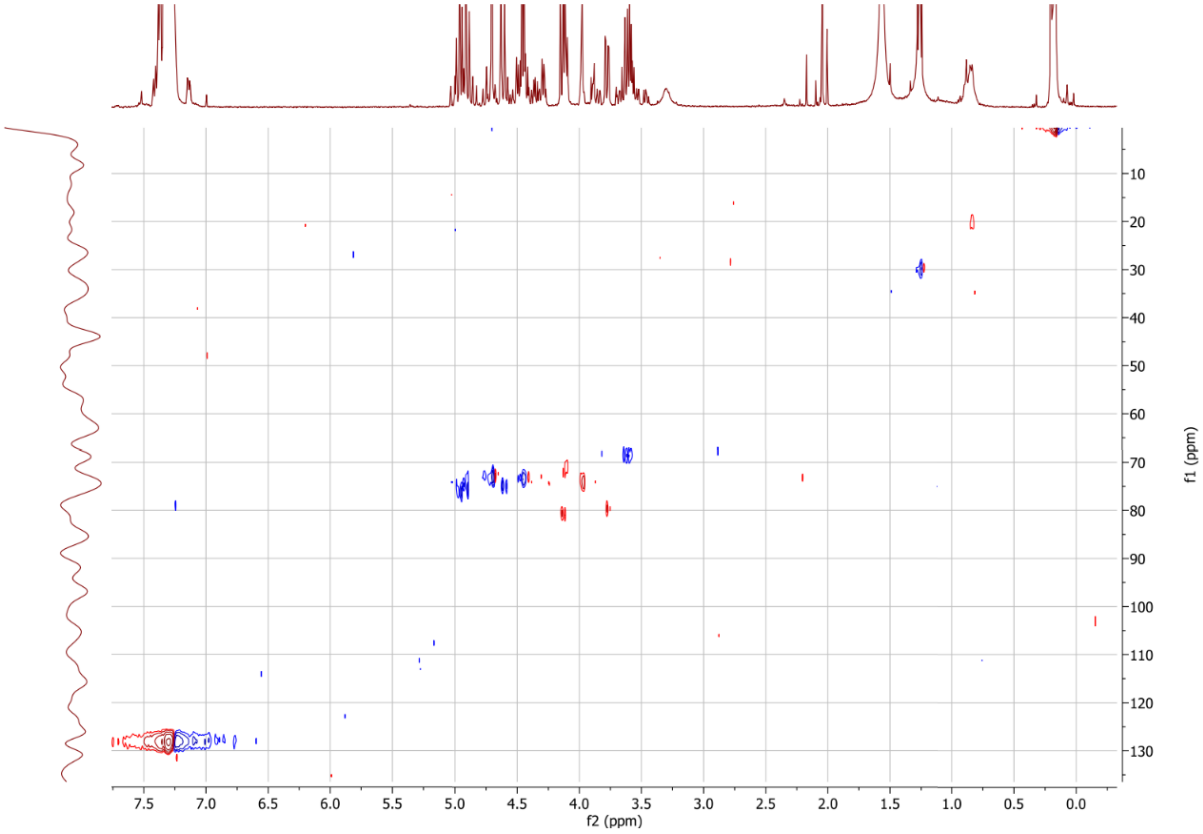
COSY:



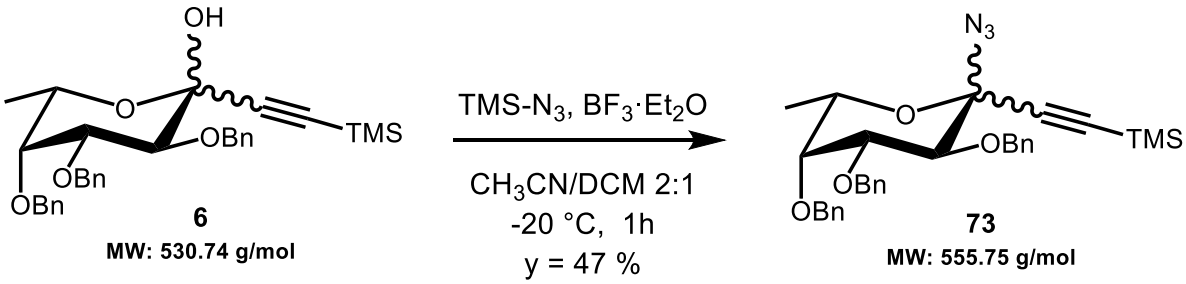
^{13}C chemical shifts were extrapolated from the HSQC experiment:

major-anomer: δ = 128.0 (*CH Ar*), 80.6 (*C2*), 79.6 (*C3*), 74.1 (*C4*), 76.2, 74.7, 73.5, 72.9 (*CH_2 Ar*), 70.8 (*C5*), 68.5, 68.4 (*C6*), 0.0 (*CH_3-Si*).

HSQC:



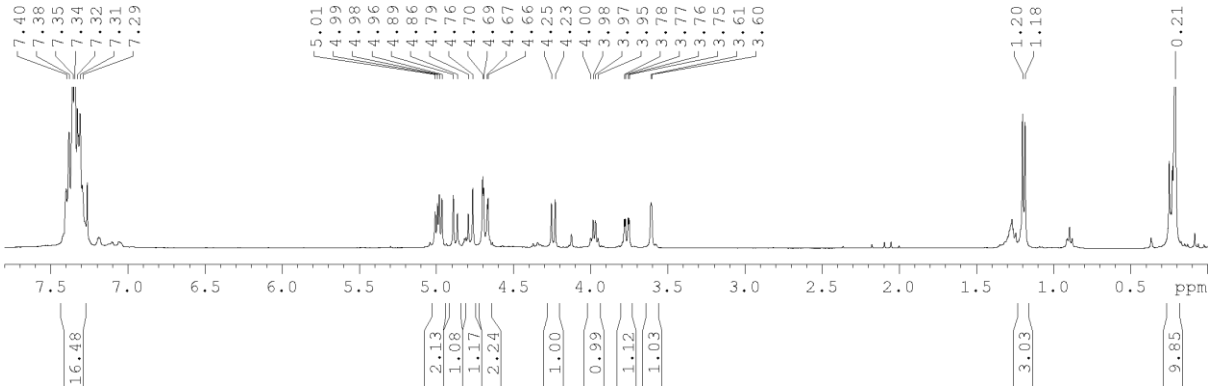
(73) Synthesis and characterization of **1-azido-2,3,4-tri-O-benzyl-1-C-trimethylsilyl acetylene-β-L-fucopyranose (73)** following the procedure of Gómez and co-workers:[203]



A solution of **1-hydroxy-2,3,4-tri-O-benzyl L-fucopyranosyl) trimethylsilyl acetylene 6** (38 mg, 0.07 mmol, 1 eq) dissolved in CH_3CN/DCM (ratio 2:1 - 1.5 mL) was cooled to -20 °C under N_2 atmosphere. TMS-N_3 (38 µL, 0.29 mmol, 4 eq), then $BF_3{\cdot}Et_2O$ (45 µL, 0.36 mmol, 5 eq) were added to the solution and left to stir at -20 °C for 20min until TLC showed no further advancement, before returning to room temperature. The reaction mixture was quenched with a few drops of Et_3N and extracted with EtOAc. The organic phase was washed with water and brine, then was dried over Na_2SO_4. The crude product was purified by flash chromatography (nHex/EtOAc: 9/1) affording product **73** (19 mg, 0.03 mmol, y = 47 %). TLC R_f (nHex/EtOAc: 7/3): 0.64.

^{1}H NMR (400 MHz, $CDCl_3$):

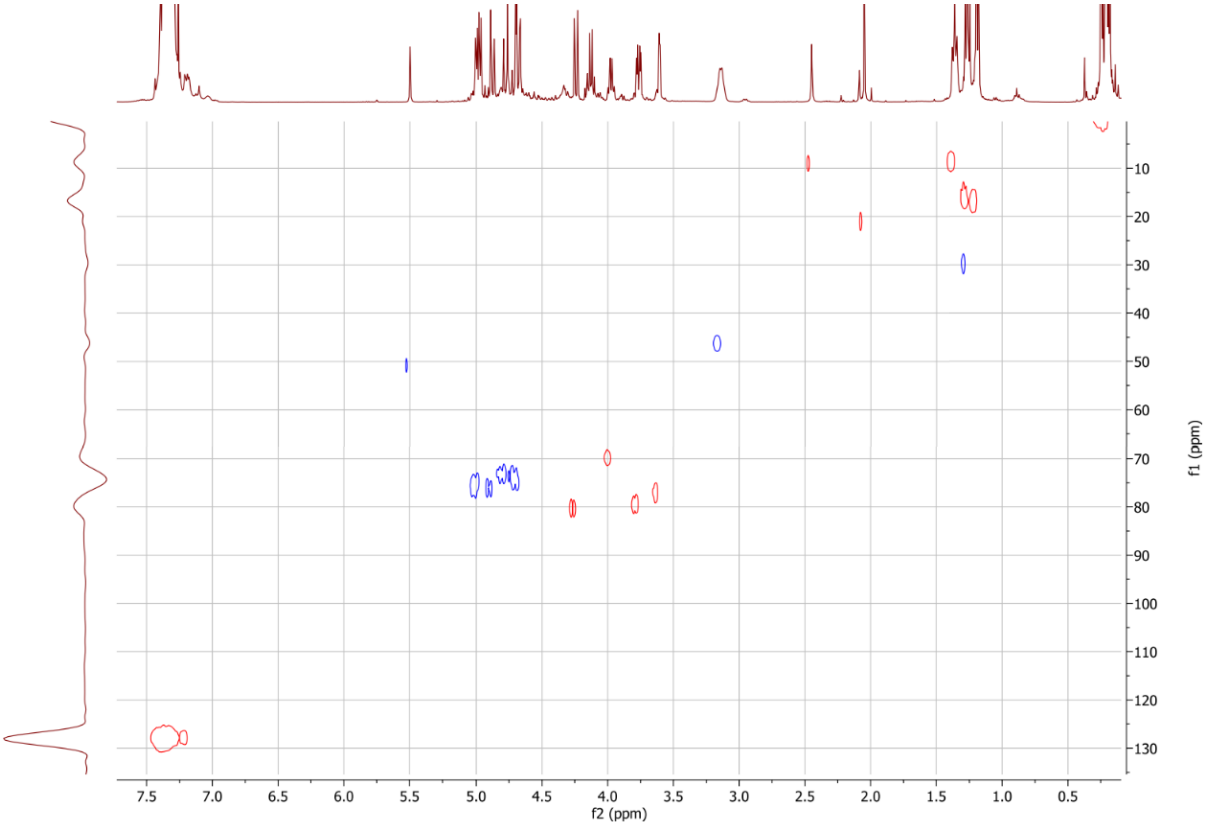
δ = 7.40 - 7.29 (m, 15H, *CH Ar*), 5.02 - 4.96 (m, 2H, *CH_2-OBn*), 4.88 (d, 1H, *CH_2-OBn*), 4.78 (d, 1H, *CH_2-OBn*), 4.70 - 4.66 (m, 2H, *CH_2-OBn*), 4.24 (d, $J_{2\text{-}3}$ = 9.8 Hz, 1H, *H-2*), 3.98 (dq, $J_{5\text{-}4}$ = 1.2 Hz, $J_{5\text{-}CH3}$ = 6.5 Hz, 1H, *H-5*), 3.77 (dd, $J_{3\text{-}2}$ = 9.8 Hz, $J_{3\text{-}4}$ = 2.8 Hz, 1H, *H-3*), 3.61 (dd, $J_{4\text{-}3}$ = 2.9 Hz, $J_{4\text{-}5}$ = 1.2 Hz, 1H, *H-4*), 1.19 (d, $J_{CH3\text{-}5}$ = 6.5 Hz, 3H, *CH_3*), 0.21 (m, 9H, *Si-CH_3*).



^{13}C shifts were extrapolated from the HSQC experiment:

δ = 127.9 (*CH Ar*), 80.3 (*C2*), 79.6 (*C3*), 77.0 (*C4*), 76.0, 74.6, 73.3 (*CH_2 Ar*), 69.8 (*C5*), 16.8 (*C6*).

HSQC:



9. REFERENCES

(1) World Health Organization. Global tuberculosis report 2020. Geneva: WHO; **2020**. (https://www.who.int/publications/i/item/9789240013131, accessed March 2021).

(2) Bushak, L. A Brief History of Antibiotic Resistance: How a Medical Miracle Turned into the Biggest Public Health Danger of Our Time. Medical Daily; **2016**. (https://www.medicaldaily.com/antibiotic-resistance-history-373773, accessed March 2021).

(3) U.S. Centers for Disease Control and Prevention. Antibiotic Resistance Threats in the United States 2019. Atlanta, GA: U.S. Department of Health and Human Services, CDC; **2019**. (https://stacks.cdc.gov/view/cdc/82532, accessed March 2021).

(4) McGann, P.; Snesrud, E.; Maybank, R.; Corey, B.; Ong, A. C. *et al. Escherichia coli* Harboring mcr-1 and blaCTX-M on a Novel IncF Plasmid: First Report of mcr-1 in the United States. *Antimicrob Agents Chemother* **2016,** *60* (7), 4420-1.

(5) Appelgren, P.; Hellstrom, I.; Weitzberg, E.; Soderlund, V.; Bindslev, L. *et al.* Risk factors for nosocomial intensive care infection: a long-term prospective analysis. *Acta Anaesthesiol Scand* **2001,** *45* (6), 710-9.

(6) Ozer, B.; Tatman-Otkun, M.; Memis, D.; Otkun, M. Nosocomial infections and risk factors in intensive care unit of a university hospital in Turkey. *Central European Journal of Medicine* **2010,** *5* (2), 203-208.

(7) van Duin, D.; Paterson, D. L. Multidrug-Resistant Bacteria in the Community: Trends and Lessons Learned. *Infect Dis Clin North Am* **2016,** *30* (2), 377-390.

(8) Cassini, A.; Hogberg, L. D.; Plachouras, D.; Quattrocchi, A.; Hoxha, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect Dis* **2019,** *19* (1), 56-66.

(9) O'Sullivan, B. P.; Freedman, S. D. Cystic fibrosis. *Lancet* **2009,** *373* (9678), 1891-904.

(10) Ciofu, O.; Hansen, C. R.; Hoiby, N. Respiratory bacterial infections in cystic fibrosis. *Curr Opin Pulm Med* **2013,** *19* (3), 251-8.

(11) European Comission. A European One Health Action Plan against Antimicrobial Resistance (AMR). Brussels: European Comission; **2017**. (https://ec.europa.eu/health/sites/health/files/antimicrobial_resistance/docs/amr_2017_action-plan.pdf, accessed March 2021).

(12) World Health Organization. Global Action Plan on Antimicrobial Resistance. Geneva: WHO; **2015**. (https://www.who.int/antimicrobial-resistance/publications/global-action-plan/en/, accessed March 2021).

(13) Carlos, T. M.; Harlan, J. M. Leukocyte-endothelial adhesion molecules. *Blood* **1994,** *84* (7), 2068-101.

(14) Cozens, D.; Read, R. C. Anti-adhesion methods as novel therapeutics for bacterial infections. *Expert Rev Anti Infect Ther* **2012,** *10* (12), 1457-68.

(15) Roberts, P. A.; Huebinger, R. M.; Keen, E.; Krachler, A. M.; Jabbari, S. Mathematical model predicts anti-adhesion-antibiotic-debridement combination therapies can clear an antibiotic resistant infection. *PLoS Comput Biol* **2019,** *15* (7), e1007211.

(16) Krachler, A. M.; Orth, K. Targeting the bacteria-host interface: strategies in anti-adhesion therapy. *Virulence* **2013,** *4* (4), 284-94.

(17) Reitsma, S.; Slaaf, D. W.; Vink, H.; van Zandvoort, M. A.; oude Egbrink, M. G. The endothelial glycocalyx: composition, functions, and visualization. *Pflugers Arch* **2007,** *454* (3), 345-59.

(18) Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C. A. Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? *Cancer Cell* **2019,** *36* (1), 6-16.

(19) Pinho, S. S.; Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015,** *15* (9), 540-555.

(20) Kaltner, H.; Abad-Rodríguez, J.; Corfield, A. P.; Kopitz, J.; Gabius, H.-J. The sugar code: letters and vocabulary, writers, editors and readers and biosignificance of functional glycan–lectin pairing. *Biochemical Journal* **2019,** *476* (18), 2623-2655.

(21) Gabius, H.-J. The sugar code: Why glycans are so important. *Biosystems* **2018,** *164*, 102-111.

(22) Solís, D.; Bovin, N. V.; Davis, A. P.; Jiménez-Barbero, J.; Romero, A. *et al.* A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015,** *1850* (1), 186-235.

(23) Zausig, Y. A.; Chappell, D.; Becker, B. F.; Potschka, D.; Busse, H. *et al.* The impact of crystalloidal and colloidal infusion preparations on coronary vascular integrity, interstitial oedema and cardiac performance in isolated hearts. *Critical Care* **2013,** *17* (5), R203.

(24) Sharon, N.; Lis, H. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* **2004,** *14* (11), 53R-62R.

(25) Notova, S.; Bonnardel, F.; Lisacek, F.; Varrot, A.; Imberty, A. Structure and engineering of tandem repeat lectins. *Current Opinion in Structural Biology* **2020,** *62*, 39-47.

(26) Ribeiro, J. P.; Villringer, S.; Goyard, D.; Coche-Guerente, L.; Höferlin, M. *et al.* Tailor-made Janus lectin with dual avidity assembles glycoconjugate multilayers and crosslinks protocells. *Chemical Science* **2018,** *9* (39), 7634-7641.

(27) Bonnardel, F.; Mariethoz, J.; Salentin, S.; Robin, X.; Schroeder, M. *et al.* UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Research* **2019,** *47* (D1), D1236-D1244.

(28) Viela, F.; Mathelie-Guinlet, M.; Viljoen, A.; Dufrene, Y. F. What makes bacterial pathogens so sticky? *Mol Microbiol* **2020,** *113* (4), 683-690.

(29) Imberty, A.; Mitchell, E. P.; Wimmerova, M. Structural basis of high-affinity glycan recognition by bacterial and fungal lectins. *Curr Opin Struct Biol* **2005,** *15* (5), 525-34.

(30) Merritt, E. A.; Hol, W. G. J. AB_5 toxins. *Current Opinion in Structural Biology* **1995,** *5* (2), 165-171.

(31) Imberty, A.; Varrot, A. Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* **2008,** *18* (5), 567-76.

(32) Poole, J.; Day, C. J.; von Itzstein, M.; Paton, J. C.; Jennings, M. P. Glycointeractions in bacterial pathogenesis. *Nature Reviews Microbiology* **2018,** *16* (7), 440-452.

(33) Hall-Stoodley, L.; Costerton, J. W.; Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* **2004,** *2* (2), 95-108.

(34) Diggle, S. P.; Stacey, R. E.; Dodd, C.; Camara, M.; Williams, P. *et al.* The galactophilic lectin, LecA, contributes to biofilm development in *Pseudomonas aeruginosa*. *Environ Microbiol* **2006,** *8* (6), 1095-104.

(35) Inhulsen, S.; Aguilar, C.; Schmid, N.; Suppiger, A.; Riedel, K. *et al.* Identification of functions linking quorum sensing with biofilm formation in *Burkholderia cenocepacia* H111. *Microbiologyopen* **2012,** *1* (2), 225-42.

(36) Tielker, D.; Hacker, S.; Loris, R.; Strathmann, M.; Wingender, J. *et al. Pseudomonas aeruginosa* lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology (Reading)* **2005,** *151* (Pt 5), 1313-1323.

(37) Heggelund, J. E.; Varrot, A.; Imberty, A.; Krengel, U. Histo-blood group antigens as mediators of infections. *Current Opinion in Structural Biology* **2017,** *44*, 190-200.

(38) Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim Biophys Acta* **2006,** *1760* (4), 527-37.

(39) Chiodo, F.; Bruijns, S. C. M.; Rodriguez, E.; Li, R. J. E.; Molinaro, A. *et al.* Novel ACE2-Independent Carbohydrate-Binding of SARS-CoV-2 Spike Protein to Host Lectins and Lung Microbiota. *bioRxiv* **2020**, 2020.05.13.092478.

(40) Letko, M.; Marzi, A.; Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **2020,** *5* (4), 562-569.

(41) Sharon, N.; Ofek, I. Safe as mother's milk: carbohydrates as future anti-adhesion drugs for bacterial diseases. *Glycoconj J* **2000,** *17* (7-9), 659-64.

(42) Aronson, M.; Medalia, O.; Schori, L.; Mirelman, D.; Sharon, N. *et al.* Prevention of colonization of the urinary tract of mice with *Escherichia coli* by blocking of bacterial adherence with methyl alpha-D-mannopyranoside. *J Infect Dis* **1979,** *139* (3), 329-32.

(43) Ukkonen, P.; Varis, K.; Jernfors, M.; Herva, E.; Jokinen, J. *et al.* Treatment of acute otitis media with an antiadhesive oligosaccharide: a randomised, double-blind, placebo-controlled trial. *Lancet* **2000,** *356* (9239), 1398-402.

(44) Parente, F.; Cucino, C.; Anderloni, A.; Grandinetti, G.; Bianchi Porro, G. Treatment of *Helicobacter pylori* infection using a novel antiadhesion compound (3'sialyllactose sodium salt). A double blind, placebo-controlled clinical study. *Helicobacter* **2003,** *8* (4), 252-6.

(45) Tamburrini, A.; Colombo, C.; Bernardi, A. Design and synthesis of glycomimetics: Recent advances. *Med Res Rev* **2020,** *40* (2), 495-531.

(46) Ernst, B.; Magnani, J. L. From carbohydrate leads to glycomimetic drugs. *Nat Rev Drug Discov* **2009,** *8* (8), 661-77.

(47) Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L. *et al.* Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J Am Chem Soc* **1997,** *119* (4), 681-90.

(48) Campbell, L. K.; Baker, D. E.; Campbell, R. K. Miglitol: assessment of its role in the treatment of patients with diabetes mellitus. *Ann Pharmacother* **2000,** *34* (11), 1291-301.

(49) Chen, X.; Zheng, Y.; Shen, Y. Voglibose (Basen, AO-128), one of the most important alpha-glucosidase inhibitors. *Curr Med Chem* **2006,** *13* (1), 109-16.

(50) World Health Organization. Global report on diabetes. Geneva: WHO; **2016**. (https://www.who.int/publications/i/item/9789241565257, Accessed March 2021).

(51) Natoni, A.; Macauley, M. S.; O'Dwyer, M. E. Targeting Selectins and Their Ligands in Cancer. *Front Oncol* **2016,** *6*, 93.

(52) Takenaka, Y.; Fukumori, T.; Raz, A. Galectin-3 and metastasis. *Glycoconj J* **2002,** *19* (7-9), 543-9.

(53) Festuccia, C.; Mancini, A.; Gravina, G. L.; Colapietro, A.; Vetuschi, A. *et al.* Dual CXCR4 and E-Selectin Inhibitor, GMI-1359, Shows Anti-Bone Metastatic Effects and Synergizes with Docetaxel in Prostate Cancer Cell Intraosseous Growth. *Cells* **2019,** *9* (1).

(54) Wdowiak, K.; Francuz, T.; Gallego-Colon, E.; Ruiz-Agamez, N.; Kubeczko, M. *et al.* Galectin Targeted Therapy in Oncology: Current Knowledge and Perspectives. *Int J Mol Sci* **2018,** *19* (1).

(55) Noble, S.; Goa, K. L. Gemcitabine. A review of its pharmacology and clinical potential in non-small cell lung cancer and pancreatic cancer. *Drugs* **1997,** *54* (3), 447-72.

(56) Hevey, R. Strategies for the Development of Glycomimetic Drug Candidates. *Pharmaceuticals (Basel)* **2019,** *12* (2).

(57) Medve, L.; Achilli, S.; Guzman-Caldentey, J.; Thepaut, M.; Senaldi, L. *et al.* Enhancing Potency and Selectivity of a DC-SIGN Glycomimetic Ligand by Fragment-Based Design: Structural Basis. *Chemistry* **2019,** *25* (64), 14659-14668.

(58) Sattin, S.; Bernardi, A. Glycoconjugates and Glycomimetics as Microbial Anti-Adhesives. *Trends Biotechnol* **2016,** *34* (6), 483-495.

(59) Ren, W.; Pengelly, R.; Farren-Dai, M.; Shamsi Kazem Abadi, S.; Oehler, V. *et al.* Revealing the mechanism for covalent inhibition of glycoside hydrolases by carbasugars at an atomic level. *Nat Commun* **2018,** *9* (1), 3243.

(60) Wagner, S.; Hauck, D.; Hoffmann, M.; Sommer, R.; Joachim, I. *et al.* Covalent Lectin Inhibition and Application in Bacterial Biofilm Imaging. *Angew Chem Int Ed Engl* **2017,** *56* (52), 16559-16564.

(61) Varga, N.; Sutkeviciute, I.; Ribeiro-Viana, R.; Berzi, A.; Ramdasi, R. *et al.* A multivalent inhibitor of the DC-SIGN dependent uptake of HIV-1 and Dengue virus. *Biomaterials* **2014,** *35* (13), 4175-4184.

(62) Martínez-Ávila, O.; Hijazi, K.; Marradi, M.; Clavel, C.; Campion, C. *et al.* Gold Manno-Glyconanoparticles: Multivalent Systems to Block HIV-1 gp120 Binding to the Lectin DC-SIGN. *Chemistry – A European Journal* **2009,** *15* (38), 9874-9888.

(63) Bernardi, A.; Jimenez-Barbero, J.; Casnati, A.; De Castro, C.; Darbre, T. *et al.* Multivalent glycoconjugates as anti-pathogenic agents. *Chem Soc Rev* **2013,** *42* (11), 4709-27.

(64) Lee, Y. C.; Lee, R. T. Carbohydrate-Protein Interactions: Basis of Glycobiology. *Accounts of Chemical Research* **1995,** *28* (8), 321-327.

(65) Lundquist, J. J.; Toone, E. J. The cluster glycoside effect. *Chem Rev* **2002,** *102* (2), 555-78.

(66) Pieters, R. J. Maximising multivalency effects in protein-carbohydrate interactions. *Org Biomol Chem* **2009,** *7* (10), 2013-25.

(67) Prost, L. R.; Grim, J. C.; Tonelli, M.; Kiessling, L. L. Noncarbohydrate glycomimetics and glycoprotein surrogates as DC-SIGN antagonists and agonists. *ACS Chem Biol* **2012,** *7* (9), 1603-8.

(68) Budhadev, D.; Poole, E.; Nehlmeier, I.; Liu, Y.; Hooper, J. *et al.* Glycan-Gold Nanoparticles as Multifunctional Probes for Multivalent Lectin-Carbohydrate Binding: Implications for Blocking Virus Infection and Nanoparticle Assembly. *J Am Chem Soc* **2020,** *142* (42), 18022-18034.

(69) Ashree, J.; Wang, Q.; Chao, Y. Glyco-functionalised quantum dots and their progress in cancer diagnosis and treatment. *Frontiers of Chemical Science and Engineering* **2020,** *14* (3), 365-377.

(70) Kim, B. S.; Hong, D. J.; Bae, J.; Lee, M. Controlled self-assembly of carbohydrate conjugate rod-coil amphiphiles for supramolecular multivalent ligands. *J Am Chem Soc* **2005,** *127* (46), 16333-7.

(71) Schaeffer, E.; Dehuyser, L.; Sigwalt, D.; Flacher, V.; Bernacchi, S. *et al.* Dynamic micelles of mannoside glycolipids are more efficient than polymers for inhibiting HIV-1 trans-infection. *Bioconjug Chem* **2013,** *24* (11), 1813-23.

(72) Soria-Martinez, L.; Bauer, S.; Giesler, M.; Schelhaas, S.; Materlik, J. *et al.* Prophylactic Antiviral Activity of Sulfated Glycomimetic Oligomers and Polymers. *J Am Chem Soc* **2020,** *142* (11), 5252-5265.

(73) Zubkova, O. V.; Ahmed, Y. A.; Guimond, S. E.; Noble, S. L.; Miller, J. H. *et al.* Dendrimer Heparan Sulfate Glycomimetics: Potent Heparanase Inhibitors for Anticancer Therapy. *ACS Chem Biol* **2018,** *13* (12), 3236-3242.

(74) Ribeiro-Viana, R.; Sanchez-Navarro, M.; Luczkowiak, J.; Koeppe, J. R.; Delgado, R. *et al.* Virus-like glycodendrinanoparticles displaying quasi-equivalent nested polyvalency upon glycoprotein platforms potently block viral infection. *Nat Commun* **2012,** *3*, 1303.

(75) Cecioni, S.; Imberty, A.; Vidal, S. Glycomimetics versus multivalent glycoconjugates for the design of high affinity lectin ligands. *Chem Rev* **2015,** *115* (1), 525-61.

(76) Kane, R. S. Thermodynamics of multivalent interactions: influence of the linker. *Langmuir* **2010,** *26* (11), 8636-40.

(77) Ordanini, S.; Varga, N.; Porkolab, V.; Thépaut, M.; Belvisi, L. *et al.* Designing nanomolar antagonists of DC-SIGN-mediated HIV infection: ligand presentation using molecular rods. *Chemical Communications* **2015,** *51* (18), 3816-3819.

(78) Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. Synthetic multivalent ligands as probes of signal transduction. *Angew Chem Int Ed Engl* **2006,** *45* (15), 2348-68.

(79) World Health Organization. Leading causes of death and disability: A visual summary of global and regional trends 2000-2019. WHO website; **2020**. (https://www.who.int/data/stories/leading-causes-of-death-and-disability-2000-2019-a-visual-summary, accessed March 2021).

(80) Whiteley, M.; Bangera, M. G.; Bumgarner, R. E.; Parsek, M. R.; Teitzel, G. M. *et al.* Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **2001,** *413* (6858), 860-4.

(81) Pimenta, A. I.; Bernardes, N.; Alves, M. M.; Mil-Homens, D.; Fialho, A. M. *Burkholderia cenocepacia* transcriptome during the early contacts with giant plasma membrane vesicles derived from live bronchial epithelial cells. *Sci Rep* **2021,** *11* (1), 5624.

(82) World Health Organization. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. Geneva: WHO; **2017**. (https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/, accessed March 2021).

(83) Diggle, S. P.; Whiteley, M. Microbe Profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat. *Microbiology (Reading)* **2020,** *166* (1), 30-33.

(84) Garcia-Clemente, M.; de la Rosa, D.; Maiz, L.; Giron, R.; Blanco, M. *et al.* Impact of P*seudomonas aeruginosa* Infection on Patients with Chronic Inflammatory Airway Diseases. *J Clin Med* **2020,** *9* (12).

(85) Gilboa-Garber, N. *Pseudomonas aeruginosa* lectins. *Methods Enzymol* **1982,** *83*, 378-85.

(86) Winzer, K.; Falconer, C.; Garber, N. C.; Diggle, S. P.; Camara, M. *et al.* The *Pseudomonas aeruginosa* lectins PA-IL and PA-IIL are controlled by quorum sensing and by RpoS. *J Bacteriol* **2000,** *182* (22), 6401-11.

(87) Passos da Silva, D.; Matwichuk, M. L.; Townsend, D. O.; Reichhardt, C.; Lamba, D. *et al.* The *Pseudomonas aeruginosa* lectin LecB binds to the exopolysaccharide Psl and stabilizes the biofilm matrix. *Nat Commun* **2019,** *10* (1), 2183.

(88) Adam, E. C.; Mitchell, B. S.; Schumacher, D. U.; Grant, G.; Schumacher, U. *Pseudomonas aeruginosa* II lectin stops human ciliary beating: therapeutic implications of fucose. *Am J Respir Crit Care Med* **1997,** *155* (6), 2102-4.

(89) Bajolet-Laudinat, O.; Girod-de Bentzmann, S.; Tournier, J. M.; Madoulet, C.; Plotkowski, M. C. *et al.* Cytotoxicity of *Pseudomonas aeruginosa* internal lectin PA-I to respiratory epithelial cells in primary culture. *Infect Immun* **1994,** *62* (10), 4481-7.

(90) Chemani, C.; Imberty, A.; de Bentzmann, S.; Pierre, M.; Wimmerova, M. *et al.* Role of LecA and LecB lectins in *Pseudomonas aeruginosa*-induced lung injury and effect of carbohydrate ligands. *Infect Immun* **2009,** *77* (5), 2065-75.

(91) Bucior, I.; Abbott, J.; Song, Y.; Matthay, M. A.; Engel, J. N. Sugar administration is an effective adjunctive therapy in the treatment of *Pseudomonas aeruginosa* pneumonia. *Am J Physiol Lung Cell Mol Physiol* **2013,** *305* (5), L352-63.

(92) Hauber, H. P.; Schulz, M.; Pforte, A.; Mack, D.; Zabel, P. *et al.* Inhalation with fucose and galactose for treatment of *Pseudomonas aeruginosa* in cystic fibrosis patients. *Int J Med Sci* **2008,** *5* (6), 371-6.

(93) Sommer, R.; Rox, K.; Wagner, S.; Hauck, D.; Henrikus, S. S. *et al.* Anti-biofilm Agents against *Pseudomonas aeruginosa*: A Structure-Activity Relationship Study of C-Glycosidic LecB Inhibitors. *J Med Chem* **2019,** *62* (20), 9201-9216.

(94) Gustke, H.; Kleene, R.; Loers, G.; Nehmann, N.; Jaehne, M. *et al.* Inhibition of the bacterial lectins of *Pseudomonas aeruginosa* with monosaccharides and peptides. *Eur J Clin Microbiol Infect Dis* **2012,** *31* (2), 207-15.

(95) Boukerb, A. M.; Rousset, A.; Galanos, N.; Mear, J. B.; Thepaut, M. *et al.* Antiadhesive properties of glycoclusters against *Pseudomonas aeruginosa* lung infection. *J Med Chem* **2014,** *57* (24), 10275-89.

(96) Sommer, R.; Wagner, S.; Rox, K.; Varrot, A.; Hauck, D. *et al.* Glycomimetic, Orally Bioavailable LecB Inhibitors Block Biofilm Formation of *Pseudomonas aeruginosa*. *J Am Chem Soc* **2018,** *140* (7), 2537-2545.

(97) Meiers, J.; Zahorska, E.; Rohrig, T.; Hauck, D.; Wagner, S. *et al.* Directing Drugs to Bugs: Antibiotic-Carbohydrate Conjugates Targeting Biofilm-Associated Lectins of *Pseudomonas aeruginosa*. *J Med Chem* **2020,** *63* (20), 11707-11724.

(98) Tavares, M.; Kozak, M.; Balola, A.; Sa-Correia, I. *Burkholderia cepacia* Complex Bacteria: a Feared Contamination Risk in Water-Based Pharmaceutical Products. *Clin Microbiol Rev* **2020,** *33* (3).

(99) Loutet, S. A.; Valvano, M. A. A decade of *Burkholderia cenocepacia* virulence determinant research. *Infect Immun* **2010,** *78* (10), 4088-100.

(100) Mil-Homens, D.; Fialho, A. M. A BCAM0223 mutant of *Burkholderia cenocepacia* is deficient in hemagglutination, serum resistance, adhesion to epithelial cells and virulence. *PLoS One* **2012,** *7* (7), e41747.

(101) Scoffone, V. C.; Chiarelli, L. R.; Trespidi, G.; Mentasti, M.; Riccardi, G. *et al. Burkholderia cenocepacia* Infections in Cystic Fibrosis Patients: Drug Resistance and Therapeutic Approaches. *Front Microbiol* **2017,** *8*, 1592.

(102) Mira, N. P.; Madeira, A.; Moreira, A. S.; Coutinho, C. P.; Sa-Correia, I. Genomic expression analysis reveals strategies of *Burkholderia cenocepacia* to adapt to cystic fibrosis patients' airways and antimicrobial therapy. *PLoS One* **2011,** *6* (12), e28831.

(103) Bragonzi, A.; Farulla, I.; Paroni, M.; Twomey, K. B.; Pirone, L. *et al.* Modelling co-infection of the cystic fibrosis lung by *Pseudomonas aeruginosa* and *Burkholderia cenocepacia* reveals influences on biofilm formation and host response. *PLoS One* **2012,** *7* (12), e52330.

(104) Speert, D. P.; Bond, M.; Woodman, R. C.; Curnutte, J. T. Infection with *Pseudomonas cepacia* in chronic granulomatous disease: role of nonoxidative killing by neutrophils in host defense. *J Infect Dis* **1994,** *170* (6), 1524-31.

(105) Mahenthiralingam, E.; Urban, T. A.; Goldberg, J. B. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat Rev Microbiol* **2005,** *3* (2), 144-56.

(106) Campana, S.; Taccetti, G.; Ravenni, N.; Favari, F.; Cariani, L. *et al.* Transmission of *Burkholderia cepacia* complex: evidence for new epidemic clones infecting cystic fibrosis patients in Italy. *J Clin Microbiol* **2005,** *43* (10), 5136-42.

(107) Vandamme, P.; Holmes, B.; Vancanneyt, M.; Coenye, T.; Hoste, B. *et al.* Occurrence of multiple genomovars of *Burkholderia cepacia* in cystic fibrosis patients and proposal of *Burkholderia multivorans* sp. nov. *Int J Syst Bacteriol* **1997,** *47* (4), 1188-200.

(108) IBCWG International *Burkholderia cepacia* Working Group. ibcwg.org (accessed March 2021).

(109) Jones, A. M.; Dodd, M. E.; Govan, J. R.; Barcus, V.; Doherty, C. J. *et al. Burkholderia cenocepacia* and *Burkholderia multivorans*: influence on survival in cystic fibrosis. *Thorax* **2004,** *59* (11), 948-51.

(110) Seth-Smith, H. M. B.; Casanova, C.; Sommerstein, R.; Meinel, D. M.; Abdelbary, M. M. H. *et al.* Phenotypic and Genomic Analyses of *Burkholderia stabilis* Clinical Contamination, Switzerland. *Emerg Infect Dis* **2019,** *25* (6), 1084-1092.

(111) ANSM Agence Nationale de Sécurité du Médicament et des produits de santé, Les désinfectants Surfa'safe premium et Opaster Anios des Laboratoires Anios ne doivent plus être utilisés - Communiqué. https://ansm.sante.fr/actualites/les-desinfectants-surfasafe-premium-et-opaster-anios-des-laboratoires-anios-ne-doivent-plus-etre-utilises-communique (accessed March 2021).

(112) Agence Nationale de Sécurité du Médicament et des produits de santé. Message d'Alerte Rapide Sanitaire - MARS N°2019_11. France: ANSM; **2019**. (https://ansm.sante.fr/uploads/2021/01/15/20191108-mars-contamination-anios-7130-144.pdf, accessed March 2021).

(113) Duthoit, B. L'entreprise nordiste Anios fait un rappel mondial de désinfectants et suspend sa production. La Voix du Nord; **2019**. (https://www.lavoixdunord.fr/662699/article/2019-11-07/l-entreprise-nordiste-anios-fait-un-rappel-mondial-de-desinfectants-et-suspend, accessed March 2021).

(114) Carlotti, A. *Burkholderia cepacia* strikes again. A3P - Association pour les Produits Propres et Parentéraux; **2020**. (https://www.a3p.org/en/burkholderia-cepacia/, accessed March 2021).

(115) Drevinek, P.; Mahenthiralingam, E. *Burkholderia cenocepacia* in cystic fibrosis: epidemiology and molecular mechanisms of virulence. *Clin Microbiol Infect* **2010,** *16* (7), 821-30.

(116) Vandamme, P.; Holmes, B.; Coenye, T.; Goris, J.; Mahenthiralingam, E. *et al. Burkholderia cenocepacia* sp. nov.--a new twist to an old story. *Res Microbiol* **2003,** *154* (2), 91-6.

(117) Holden, M. T.; Seth-Smith, H. M.; Crossman, L. C.; Sebaihia, M.; Bentley, S. D. *et al.* The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. *J Bacteriol* **2009,** *191* (1), 261-77.

(118) Regan, K. H.; Bhatt, J. Eradication therapy for *Burkholderia cepacia* complex in people with cystic fibrosis. *Cochrane Database Syst Rev* **2019,** *4*, CD009876.

(119) Mok, B. Y.; de Moraes, M. H.; Zeng, J.; Bosch, D. E.; Kotrys, A. V. *et al.* A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **2020,** *583* (7817), 631-637.

(120) Lameignere, E.; Malinovska, L.; Slavikova, M.; Duchaud, E.; Mitchell, E. P. *et al.* Structural basis for mannose recognition by a lectin from opportunistic bacteria *Burkholderia cenocepacia*. *Biochem J* **2008,** *411* (2), 307-18.

(121) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N. *et al.* The Protein Data Bank. *Nucleic Acids Res* **2000,** *28* (1), 235-42.

(122) RCSB PDB, Research Collaboratory for Structural Bioinformatics - Protein Data Bank. https://www.rcsb.org/ (accessed March 2021).

(123) Mitchell, E.; Houles, C.; Sudakevitz, D.; Wimmerova, M.; Gautier, C. *et al.* Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nat Struct Biol* **2002,** *9* (12), 918-21.

(124) Sulak, O.; Cioci, G.; Lameignere, E.; Balloy, V.; Round, A. *et al. Burkholderia cenocepacia* BC2L-C is a super lectin with dual specificity and proinflammatory activity. *PLoS Pathog* **2011,** *7* (9), e1002238.

(125) Cory, S.; Adams, J. M. The Bcl2 family: regulators of the cellular life-or-death switch. *Nat Rev Cancer* **2002,** *2* (9), 647-56.

(126) Sabin, C.; Mitchell, E. P.; Pokorna, M.; Gautier, C.; Utille, J. P. *et al.* Binding of different monosaccharides by lectin PA-IIL from *Pseudomonas aeruginosa*: thermodynamics data correlated with X-ray structures. *FEBS Lett* **2006,** *580* (3), 982-7.

(127) Lameignere, E.; Shiao, T. C.; Roy, R.; Wimmerova, M.; Dubreuil, F. *et al.* Structural basis of the affinity for oligomannosides and analogs displayed by BC2L-A, a *Burkholderia cenocepacia* soluble lectin. *Glycobiology* **2010,** *20* (1), 87-98.

(128) Beshr, G.; Sommer, R.; Hauck, D.; Siebert, D. C. B.; Hofmann, A. *et al.* Development of a competitive binding assay for the *Burkholderia cenocepacia* lectin BC2L-A and structure activity relationship of natural and synthetic inhibitors. *MedChemComm* **2016,** *7* (3), 519-530.

(129) Csavas, M.; Malinovska, L.; Perret, F.; Gyurko, M.; Illyes, Z. T. *et al.* Tri- and tetravalent mannoclusters cross-link and aggregate BC2L-A lectin from *Burkholderia cenocepacia*. *Carbohydr Res* **2017,** *437*, 1-8.

(130) Pifferi, C.; Goyard, D.; Gillon, E.; Imberty, A.; Renaudet, O. Synthesis of Mannosylated Glycodendrimers and Evaluation against BC2L-A Lectin from *Burkholderia Cenocepacia*. *Chempluschem* **2017,** *82* (3), 390-398.

(131) Reynolds, M.; Marradi, M.; Imberty, A.; Penades, S.; Perez, S. Influence of ligand presentation density on the molecular recognition of mannose-functionalised glyconanoparticles by bacterial lectin BC2L-A. *Glycoconj J* **2013,** *30* (8), 747-57.

(132) Marchetti, R.; Malinovska, L.; Lameignere, E.; Adamova, L.; de Castro, C. *et al. Burkholderia cenocepacia* lectin A binding to heptoses from the bacterial lipopolysaccharide. *Glycobiology* **2012,** *22* (10), 1387-98.

(133) Schmid, N.; Pessi, G.; Deng, Y.; Aguilar, C.; Carlier, A. L. *et al.* The AHL- and BDSF-dependent quorum sensing systems control specific and overlapping sets of genes in *Burkholderia cenocepacia* H111. *PLoS One* **2012,** *7* (11), e49966.

(134) Sulak, O.; Cioci, G.; Delia, M.; Lahmann, M.; Varrot, A. *et al.* A TNF-like trimeric lectin domain from *Burkholderia cenocepacia* with specificity for fucosylated human histo-blood group antigens. *Structure* **2010,** *18* (1), 59-72.

(135) Houser, J.; Kosourova, J.; Kubickova, M.; Wimmerova, M. Development of 48-condition buffer screen for protein stability assessment. *Eur Biophys J* **2021**.

(136) Sulak, O.; Cioci, G.; Lameignere, E.; Delia, M.; Wimmerova, M.; Imberty, A. **2011,** PDB: 2XR4 C-terminal domain of BC2L-C Lectin from *Burkholderia cenocepacia*. DOI: 10.2210/pdb2XR4/pdb.

(137) Tateno, H.; Toyota, M.; Saito, S.; Onuma, Y.; Ito, Y. *et al.* Glycome diagnosis of human induced pluripotent stem cells using lectin microarray. *J Biol Chem* **2011,** *286* (23), 20345-53.

(138) Onuma, Y.; Tateno, H.; Hirabayashi, J.; Ito, Y.; Asashima, M. rBC2LCN, a new probe for live cell imaging of human pluripotent stem cells. *Biochem Biophys Res Commun* **2013,** *431* (3), 524-9.

(139) Tateno, H.; Hiemori, K.; Minoshima, F.; Kiyoi, K.; Matoba, K. *et al.* Oriented immobilization of rBC2LCN lectin for highly sensitive detection of human pluripotent stem cells using cell culture supernatants. *J Biosci Bioeng* **2020,** *129* (2), 215-222.

(140) Tateno, H.; Matsushima, A.; Hiemori, K.; Onuma, Y.; Ito, Y. *et al.* Podocalyxin is a glycoprotein ligand of the human pluripotent stem cell-specific probe rBC2LCN. *Stem Cells Transl Med* **2013,** *2* (4), 265-73.

(141) Tateno, H.; Onuma, Y.; Ito, Y.; Hiemori, K.; Aiki, Y. *et al.* A medium hyperglycosylated podocalyxin enables noninvasive and quantitative detection of tumorigenic human pluripotent stem cells. *Sci Rep* **2014,** *4*, 4069.

(142) Tateno, H.; Onuma, Y.; Ito, Y.; Minoshima, F.; Saito, S. *et al.* Elimination of tumorigenic human pluripotent stem cells by a recombinant lectin-toxin fusion protein. *Stem Cell Reports* **2015,** *4* (5), 811-20.

(143) Shimomura, O.; Oda, T.; Tateno, H.; Ozawa, Y.; Kimura, S. *et al.* A Novel Therapeutic Strategy for Pancreatic Cancer: Targeting Cell Surface Glycan Using rBC2LC-N Lectin-Drug Conjugate (LDC). *Mol Cancer Ther* **2018,** *17* (1), 183-195.

(144) Mawaribuchi, S.; Onuma, Y.; Aiki, Y.; Kuriyama, Y.; Mutoh, M. *et al.* The rBC2LCN-positive subpopulation of PC-3cells exhibits cancer stem-like properties. *Biochem Biophys Res Commun* **2019,** *515* (1), 176-182.

(145) Mawaribuchi, S.; Haramoto, Y.; Tateno, H.; Onuma, Y.; Aiki, Y. *et al.* rBC2LCN lectin as a potential probe of early-stage HER2-positive breast carcinoma. *FEBS Open Bio* **2020,** *10* (6), 1056-1064.

(146) Breiman, A.; Lopez Robles, M. D.; de Carne Trecesson, S.; Echasserieau, K.; Bernardeau, K. *et al.* Carcinoma-associated fucosylated antigens are markers of the epithelial state and can contribute to cell adhesion through CLEC17A (Prolectin). *Oncotarget* **2016,** *7* (12), 14064-82.

(147) Sugahara, D.; Kobayashi, Y.; Akimoto, Y.; Kawakami, H. Mouse intestinal niche cells express a distinct alpha1,2-fucosylated glycan recognized by a lectin from *Burkholderia cenocepacia*. *Glycobiology* **2017,** *27* (3), 246-253.

(148) Ziganshina, M. M.; Kulikova, G. V.; Fayzullina, N. M.; Yarotskaya, E. L.; Shchegolev, A. I. *et al.* Expression of fucosylated glycans in endothelial glycocalyces of placental villi at early and late fetal growth restriction. *Placenta* **2020,** *90*, 98-102.

(149) Geissner, A.; Reinhardt, A.; Rademacher, C.; Johannssen, T.; Monteiro, J. *et al.* Microbe-focused glycan array screening platform. *Proc Natl Acad Sci U S A* **2019,** *116* (6), 1958-1967.

(150) Tokiwa, T.; Nakano, S.; Yamamoto, Y.; Ishikawa, T.; Ito, S. *et al.* Development of an Analysis Toolkit, AnalysisFMO, to Visualize Interaction Energies Generated by Fragment Molecular Orbital Calculations. *J Chem Inf Model* **2019,** *59* (1), 25-30.

(151) Kasakova, M.; Malinovska, L.; Klejch, T.; Hlavackova, M.; Dvorakova, H. *et al.* Selectivity of original C-hexopyranosyl calix[4]arene conjugates towards lectins of different origin. *Carbohydr Res* **2018,** *469*, 60-72.

(152) Thai Le, S.; Malinovska, L.; Vaskova, M.; Mezo, E.; Kelemen, V. *et al.* Investigation of the Binding Affinity of a Broad Array of l-Fucosides with Six Fucose-Specific Lectins of Bacterial and Fungal Origin. *Molecules* **2019,** *24* (12).

(153) Qing, G.; Ma, L. C.; Khorchid, A.; Swapna, G. V.; Mal, T. K. *et al.* Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat Biotechnol* **2004,** *22* (7), 877-82.

(154) Kapust, R. B.; Tozser, J.; Fox, J. D.; Anderson, D. E.; Cherry, S. *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **2001,** *14* (12), 993-1000.

(155) Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 125-32.

(156) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 235-42.

(157) Menéndez, M. Isothermal Titration Calorimetry: Principles and Applications. *eLS* **2020**, 113-127.

(158) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **1989,** *179* (1), 131-7.

(159) Zeng, S.; Baillargeat, D.; Ho, H. P.; Yong, K. T. Nanomaterials enhanced surface plasmon resonance for biological and chemical sensing applications. *Chem Soc Rev* **2014,** *43* (10), 3426-52.

(160) de Mol, N. J. Affinity Constants for Small Molecules from SPR Competition Experiments. In Surface Plasmon Resonance: Methods and Protocols, ed.; N. J. Mol and M. J. E. Fischer, Ed.^Eds. Humana Press: Totowa, NJ, 2010; Vol. p^pp 101-111.

(161) Vasile, F.; Della Volpe, S.; Ambrosio, F. A.; Costa, G.; Unver, M. Y. *et al.* Exploration of ligand binding modes towards the identification of compounds targeting HuR: a combined STD-NMR and Molecular Modelling approach. *Sci Rep* **2018,** *8* (1), 13780.

(162) PhD4GlycoDrug Marie Skłodowska-Curie Innovative Training Network. https://www.phd4glycodrug.eu/ (accessed March 2021).

(163) Glycopedia Glycoscience Portal. https://www.glycopedia.eu/ (accessed March 2021).

(164) Lal, K.; Bermeo, R.; Perez, S. Computational tools for drawing, building and displaying carbohydrates: a visual guide. *Beilstein J Org Chem* **2020,** *16*, 2448-2468.

(165) Bermeo, R.; Bernardi, A.; Varrot, A. BC2L-C N-Terminal Lectin Domain Complexed with Histo Blood Group Oligosaccharides Provides New Structural Information. *Molecules* **2020,** *25* (2).

(166) Cramer, J.; Sager, C. P.; Ernst, B. Hydroxyl Groups in Synthetic and Natural-Product-Derived Therapeutics: A Perspective on a Common Functional Group. *Journal of Medicinal Chemistry* **2019,** *62* (20), 8915-8930.

(167) Lal, K.; Bermeo, R.; Cramer, J.; Vasile, F.; Ernst, B. *et al.* Prediction and Validation of a Druggable Site on Virulence Factor of Drug Resistant *Burkholderia cenocepacia*. *Chemistry – A European Journal* **2021,** *In press. https://doi.org/10.1002/chem.202100252*.

(168) Navarra, G.; Zihlmann, P.; Jakob, R. P.; Stangier, K.; Preston, R. C. *et al.* Carbohydrate-Lectin Interactions: An Unexpected Contribution to Affinity. *Chembiochem* **2017,** *18* (6), 539-544.

(169) Dingjan, T.; Gillon, É.; Imberty, A.; Pérez, S.; Titz, A. *et al.* Virtual Screening Against Carbohydrate-Binding Proteins: Evaluation and Application to Bacterial *Burkholderia ambifaria* Lectin. *Journal of Chemical Information and Modeling* **2018,** *58* (9), 1976-1989.

(170) Sulak, O.; Cioci, G.; Lahman, M.; Delia, M.; Varrot, A.; Imberty, A.; Wimmerova, M. **2010,** PDB: 2WQ4 N-terminal domain of BC2L-C Lectin from *Burkholderia cenocepacia*. DOI: 10.2210/pdb2WQ4/pdb.

(171) Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* **2009,** *49* (2), 377-89.

(172) Weis, W. I.; Drickamer, K. STRUCTURAL BASIS OF LECTIN-CARBOHYDRATE RECOGNITION. *Annual Review of Biochemistry* **1996,** *65* (1), 441-473.

(173) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004,** *47* (7), 1739-1749.

(174) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie International Edition* **2001,** *40* (11), 2004-2021.

(175) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angewandte Chemie International Edition* **2002,** *41* (14), 2596-2599.

(176) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *The Journal of Organic Chemistry* **2002,** *67* (9), 3057-3064.

(177) Sonogashira, K.; Tohda, Y.; Hagihara, N. A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Letters* **1975,** *16* (50), 4467-4470.

(178) Heck, R. F. Acylation, methylation, and carboxyalkylation of olefins by Group VIII metal derivatives. *Journal of the American Chemical Society* **1968,** *90* (20), 5518-5526.

(179) Lancelin, J.-M.; Zollo, P. H. A.; Sinaÿ, P. Synthesis and conversions of C-(alkyn-1-yl)-β-D-glucopyranosides. *Tetrahedron Letters* **1983,** *24* (44), 4833-4836.

(180) Alzeer, J.; Vasella, A. Oligosaccharide Analogues of Polysaccharides. Part 2. Regioselective deprotection of monosaccharide-derived monomers and dimers. *Helvetica Chimica Acta* **1995,** *78* (1), 177-193.

(181) Lowary, T.; Meldal, M.; Helmboldt, A.; Vasella, A.; Bock, K. Novel Type of Rigid C-Linked Glycosylacetylene−Phenylalanine Building Blocks for Combinatorial Synthesis of C-linked Glycopeptides. *The Journal of Organic Chemistry* **1998,** *63* (26), 9657-9668.

(182) Mowery, D. F. Isomer distribution during methyl fucoside, formation by the fischer method. Further support for the previously proposed reaction mechanism. *Carbohydrate Research* **1975,** *43* (2), 233-238.

(183) Koch, S.; Schollmeyer, D.; Löwe, H.; Kunz, H. C-Glycosyl Amino Acids through Hydroboration–Cross-Coupling of exo-Glycals and Their Application in Automated Solid-Phase Synthesis. *Chemistry – A European Journal* **2013,** *19* (22), 7020-7041.

(184) Wei, P.; Zhang, D.; Gao, Z.; Cai, W.; Xu, W. *et al.* Iodine Monochloride (ICl) as a Highly Efficient, Green Oxidant for the Oxidation of Alcohols to Corresponding Carbonyl Compounds. *Synthetic Communications* **2015,** *45* (12), 1457-1470.

(185) Adlington, M. G.; Orfanopoulos, M.; Fry, J. L. A convenient one-step synthesis of hydrocarbons from alcohols through use of the organosilane-boron trifluoride reducing system. *Tetrahedron Letters* **1976,** *17* (34), 2955-2958.

(186) Deslongchamps, P.; Rowan, D. D.; Pothier, N. The acid-catalyzed oxido-reduction of spiroketals. Evidence for stereoelectronic control in hydride transfer to cyclic oxenium ions. *Canadian Journal of Chemistry* **1981,** *59* (18), 2787-2802.

(187) Lewis, M. D.; Cha, J. K.; Kishi, Y. Highly stereoselective approaches to .alpha.- and .beta.-C-glycopyranosides. *Journal of the American Chemical Society* **1982,** *104* (18), 4976-4978.

(188) Terauchi, M.; Abe, H.; Matsuda, A.; Shuto, S. An Efficient Synthesis of β-C-Glycosides Based on the Conformational Restriction Strategy: Lewis Acid Promoted Silane Reduction of the Anomeric Position with Complete Stereoselectivity. *Organic Letters* **2004,** *6* (21), 3751-3754.

(189) Balmond, E. I.; Benito-Alifonso, D.; Coe, D. M.; Alder, R. W.; McGarrigle, E. M. *et al.* A 3,4-trans-Fused Cyclic Protecting Group Facilitates α-Selective Catalytic Synthesis of 2-Deoxyglycosides. *Angewandte Chemie International Edition* **2014,** *53* (31), 8190-8194.

(190) Rouzier, F.; Sillé, R.; Nourry, A.; Tessier, A.; Pipelier, M. *et al.* Practical Gram-Scale Synthesis of Either α- or β-Anomer of C-Vinyl Glycosides. *Synthesis* **2019,** *51* (12), 2484-2488.

(191) Kondor, Z.; Herczeg, M.; Borbás, A.; Patonay, T.; Kónya, K. Application of Carbohydrates with Methylene or Vinyl Groups in Heck–Mizoroki Cross-Coupling Reactions with O-Heterocycles. *Synlett* **2016,** *27* (19), 2709-2715.

(192) Dondoni, A.; Mariotti, G.; Marra, A. Synthesis of α- and β-Glycosyl Asparagine Ethylene Isosteres (C-Glycosyl Asparagines) via Sugar Acetylenes and Garner Aldehyde Coupling. *The Journal of Organic Chemistry* **2002,** *67* (13), 4475-4486.

(193) Frantz, D. E.; Fässler, R.; Carreira, E. M. Facile Enantioselective Synthesis of Propargylic Alcohols by Direct Addition of Terminal Alkynes to Aldehydes. *Journal of the American Chemical Society* **2000,** *122* (8), 1806-1807.

(194) Boyall, D.; Frantz, D. E.; Carreira, E. M. Efficient Enantioselective Additions of Terminal Alkynes and Aldehydes under Operationally Convenient Conditions. *Organic Letters* **2002,** *4* (15), 2605-2606.

(195) Bolchi, C.; Valoti, E.; Straniero, V.; Ruggeri, P.; Pallavicini, M. From 2-Aminomethyl-1,4-benzodioxane Enantiomers to Unichiral 2-Cyano- and 2-Carbonyl-Substituted Benzodioxanes via Dichloroamine. *The Journal of Organic Chemistry* **2014,** *79* (14), 6732-6737.

(196) Györgydeák, Z.; Szilágyi, L.; Paulsen, H. Synthesis, Structure and Reactions of Glycosyl Azides. *Journal of Carbohydrate Chemistry* **1993,** *12* (2), 139-163.

(197) Kunz, H.; Pfrengle, W.; Rück, K.; Sager, W. Stereoselective Synthesis of L-Amino Acids via Strecker and Ugi Reaktions on Carbohydrate Templates. *Synthesis* **1991,** *1991* (11), 1039-1042.

(198) Tanaka, T.; Nagai, H.; Noguchi, M.; Kobayashi, A.; Shoda, S.-i. One-step conversion of unprotected sugars to β-glycosyl azides using 2-chloroimidazolinium salt in aqueous solution. *Chemical Communications* **2009,** (23), 3378-3379.

(199) Vinson, N.; Gou, Y.; Becer, C. R.; Haddleton, D. M.; Gibson, M. I. Optimised 'click' synthesis of glycopolymers with mono/di- and trisaccharides. *Polymer Chemistry* **2011,** *2* (1), 107-113.

(200) Doores, K. J.; Mimura, Y.; Dwek, R. A.; Rudd, P. M.; Elliott, T. *et al.* Direct deprotected glycosyl–asparagine ligation. *Chemical Communications* **2006,** (13), 1401-1403.

(201) Huang, H.; Liu, H.; Jiang, H.; Chen, K. Rapid and Efficient Pd-Catalyzed Sonogashira Coupling of Aryl Chlorides. *The Journal of Organic Chemistry* **2008,** *73* (15), 6037-6040.

(202) Rico, L.; Hanessian, S. Synthesis of 1',2'-methano-2',3'-dideoxynucleosides as potential antivirals. *Bioorganic & Medicinal Chemistry Letters* **2019,** *29* (4), 597-600.

(203) Gómez, Ana M.; Uriel, C.; Jarosz, S.; Valverde, S.; López, J. C. Stereoselective Synthesis of C- and N-Ketosides by Lewis Acid-Catalyzed C- and N-Glycosidation of Alkynyl, Phenyl, and Methyl Ketoses. *European Journal of Organic Chemistry* **2003,** *2003* (24), 4830-4837.

(204) Gómez, A. M.; Uriel, C.; Valverde, S.; López, J. C. Formation and Reactivity of Novel Pyranosidic Nicholas Oxocarbenium Ions: Access to C-Ketosides and Branched-Chain C-Ketosides. *Organic Letters* **2006,** *8* (15), 3187-3190.

(205) Groothuys, S.; van den Broek, S. A. M. W.; Kuijpers, B. H. M.; Ijsselstijn, M.; van Delft, F. L. *et al.* Ring-Closing Alkyne Metathesis in the Synthesis of Alkyne-Linked Glycoamino Acids. *Synlett* **2008,** *2008* (01), 111-115.

(206) Audfray, A.; Claudinon, J.; Abounit, S.; Ruvoën-Clouet, N.; Larson, G. *et al.* Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes. *The Journal of biological chemistry* **2012,** *287* (6), 4335-4347.

(207) Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W.; Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J Virol* **2007,** *81* (13), 6807-16.

(208) Legrand, P. XDSME: XDS Made Easier. *GitHub Repos.* **2017**.

(209) McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **2007,** *63* (Pt 1), 32-41.

(210) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 486-501.

(211) Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **2011,** *67* (Pt 4), 355-67.

(212) Agirre, J.; Iglesias-Fernandez, J.; Rovira, C.; Davies, G. J.; Wilson, K. S. *et al.* Privateer: software for the conformational validation of carbohydrate structures. *Nat Struct Mol Biol* **2015,** *22* (11), 833-4.

(213) Wang, Q.; Grkovic, T.; Font, J.; Bonham, S.; Pouwer, R. H. *et al.* Monoterpene Glycoside ESK246 from Pittosporum Targets LAT3 Amino Acid Transport and Prostate Cancer Cell Growth. *ACS Chemical Biology* **2014,** *9* (6), 1369-1376.

(214) Jansson, P.-E.; Kenne, L.; Widmalm, G. Computer-assisted structural analysis of oligosaccharides using CASPER. *Analytical Biochemistry* **1991,** *199* (1), 11-17.

(215) Khaled, A.; Piotrowska, O.; Dominiak, K.; Augé, C. Exploring specificity of glycosyltransferases: synthesis of new sugar nucleotide related molecules as putative donor substrates. *Carbohydrate Research* **2008,** *343* (2), 167-178.

(216) Nishi, Y.; Tanimoto, T. Preparation and Characterization of Branched β-Cyclodextrins Having α-l-Fucopyranose and a Study of Their Functions. *Bioscience, Biotechnology, and Biochemistry* **2009,** *73* (3), 562-569.

(217) Matwiejuk, M.; Thiem, J. New Method for Regioselective Glycosylation Employing Saccharide Oxyanions. *European Journal of Organic Chemistry* **2011,** *2011* (29), 5860-5878.

(218) Frédéric, C. J. M.; Tikad, A.; Fu, J.; Pan, W.; Zheng, R. B. *et al.* Synthesis of Unprecedented Sulfonylated Phosphono-exo-Glycals Designed as Inhibitors of the Three Mycobacterial Galactofuranose Processing Enzymes. *Chemistry – A European Journal* **2016,** *22* (44), 15913-15920.

(219) Fusaro, M. B.; Chagnault, V.; Josse, S.; Postel, D. Metal-free oxidative lactonization of carbohydrates using molecular iodine. *Tetrahedron* **2013,** *69* (29), 5880-5883.

(220) Duléry, V.; Renaudet, O.; Philouze, C.; Dumy, P. α and βl-Fucopyranosyl oxyamines: key intermediates for the preparation of fucose-containing glycoconjugates by oxime ligation. *Carbohydrate Research* **2007,** *342* (7), 894-900.

(221) Machida, T.; Lang, K.; Xue, L.; Chin, J. W.; Winssinger, N. Site-Specific Glycoconjugation of Protein via Bioorthogonal Tetrazine Cycloaddition with a Genetically Encoded trans-Cyclooctene or Bicyclononyne. *Bioconjugate Chemistry* **2015,** *26* (5), 802-806.

(222) Palomo, C.; Aizpurua, J. M.; Balentová, E.; Azcune, I.; Santos, J. I. *et al.* "Click" Saccharide/β-Lactam Hybrids for Lectin Inhibition. *Organic Letters* **2008,** *10* (11), 2227-2230.

(223) George, S.; Suryavanshi, G. S.; Sudalai, A. A short enantioselective synthesis of (−)-bestatin via l-proline-catalyzed α-amination of an aldehyde. *Tetrahedron Letters* **2008,** *49* (48), 6791-6793.

(224) Baccon-Sollier, P. L.; Malki, Y.; Maye, M.; Ali, L. M. A.; Lichon, L. *et al.* Imidazopyridine-fused [1,3]diazepinones: modulations of positions 2 to 4 and their impacts on the anti-melanoma activity. *Journal of Enzyme Inhibition and Medicinal Chemistry* **2020,** *35* (1), 935-949.

(225) Vilaivan, T. A rate enhancement of tert-butoxycarbonylation of aromatic amines with Boc2O in alcoholic solvents. *Tetrahedron Letters* **2006,** *47* (38), 6739-6742.

(226) Reggelin, M.; Junker, B.; Heinrich, T.; Slavik, S.; Bühle, P. Asymmetric Synthesis of Highly Substituted Azapolycyclic Compounds via 2-Alkenyl Sulfoximines: Potential Scaffolds for Peptide Mimetics. *Journal of the American Chemical Society* **2006,** *128* (12), 4023-4034.

(227) Lei, H.; Stoakes, M. S.; Schwabacher, A. W.; Herath, K. P. B.; Lee, J. Efficient Synthesis of a Phosphinate Bis-Amino Acid and Its Use in the Construction of Amphiphilic Peptides. *The Journal of Organic Chemistry* **1994,** *59* (15), 4206-4210.

(228) Newton, A. S.; Deiana, L.; Puleo, D. E.; Cisneros, J. A.; Cutrona, K. J. *et al.* JAK2 JH2 Fluorescence Polarization Assay and Crystal Structures for Complexes with Three Small Molecules. *ACS Medicinal Chemistry Letters* **2017,** *8* (6), 614-617.

(229) Cinelli, M. A.; Reidl, C. T.; Li, H.; Chreifi, G.; Poulos, T. L. *et al.* First Contact: 7-Phenyl-2-Aminoquinolines, Potent and Selective Neuronal Nitric Oxide Synthase Inhibitors That Target an Isoform-Specific Aspartate. *Journal of Medicinal Chemistry* **2020,** *63* (9), 4528-4554.

(230) Decréau, R. A.; Collman, J. P.; Yang, Y.; Yan, Y.; Devaraj, N. K. Syntheses of hemoprotein models that can be covalently attached onto electrode surfaces by click chemistry. *The Journal of organic chemistry* **2007,** *72* (8), 2794-2802.

(231) Dubost, E.; Le Nouën, D.; Streith, J.; Tarnus, C.; Tschamber, T. Synthesis of Substituted Imidazolo[1,2-a]piperidinoses and Their Evaluation as Glycosidase Inhibitors. *European Journal of Organic Chemistry* **2006,** *2006* (3), 610-626.

(232) Bianchi, A.; Russo, A.; Bernardi, A. Neo-glycoconjugates: stereoselective synthesis of α-glycosyl amides via Staudinger ligation reactions. *Tetrahedron: Asymmetry* **2005,** *16* (2), 381-386.

(233) Hribernik, N.; Tamburrini, A.; Falletta, E.; Bernardi, A. One pot synthesis of thio-glycosides via aziridine opening reactions. *Organic & Biomolecular Chemistry* **2021,** *19* (1), 233-247.

(234) Dedola, S.; Hughes, D. L.; Nepogodiev, S. A.; Rejzek, M.; Field, R. A. Synthesis of α- and β-d-glucopyranosyl triazoles by CuAAC 'click chemistry': reactant tolerance, reaction rate, product structure and glucosidase inhibitory properties. *Carbohydrate Research* **2010,** *345* (9), 1123-1134.

(235) Pater, R. H.; Coelho, R. A.; Mowery, D. F. Chromatographic adsorption. VI. Isomer distribution and mechanism of formation of the methyl glycosides of D-glucose and D-galactose by the Fischer method. *The Journal of Organic Chemistry* **1973,** *38* (19), 3272-3277.

(236) Shi, Y.-H.; Xie, Y.-F.; Liu, Y.-Q.; Wei, Q.-C.; Xu, W.-R. *et al.* SrCl2 as an efficient cocatalyst for acidic hydrolysis of methyl glycosides. *Chinese Chemical Letters* **2014,** *25* (4), 561-566.

Conception, synthèse et évaluation de glycocomposés dirigés contre BC2L-C

Ce projet vise à contrer pour la première fois la superlectine BC2L-C appartenant à la bactérie *Burkholderia cenocepacia*, multirésistante aux médicaments.

Les pathogènes résistants tels que Burkholderia cenocepacia représentent un grave danger dans le contexte des infections nosocomiales, en particulier pour les patients affectés par la mucoviscidose ou à déficit immunitaire. Comme d'autres bactéries opportunistes à Gram négatif, ce pathogène établit sa virulence et biofilms par adhésion à travers de lectines. En particulier, la superlectine BC2L-C est soupçonnée d'être un point de rattachement entre cellules bactériennes de B. cenocepacia et cellules épithéliales humaines au cours de l'infection pulmonaire.

Dans le but d'inhiber l'extrémité N-terminale de BC2L-C, qui cible des oligosaccharides humains, nous visons à concevoir des antagonistes glycomimétiques. Nous rapportons l'étude structurale de la cible BC2L-C-N-ter par cristallographie de rayons X, suivie par la conception et synthèse d'une bibliothèque modulaire de glycomimétiques : C- et N-fucosides. Enfin, nous rapportons l'évaluation biophysique des interactions entre les glycomimétiques générés et BC2L-CN_{ter} par les techniques STD NMR, SPR, ITC, DSC; résultant en un composé principal avec une affinité satisfaisante et deux structures cristallines de complexes antagoniste/lectine.

Design, synthesis and evaluation of antagonists towards BC2L-C

This project aims to antagonize for the first time the superlectin BC2L-C from multi-drug resistant (MDR) pathogen *Burkholderia cenocepacia*.

MDRs such as Burkholderia cenocepacia have become a hazard in the context of healthcare-associated infections, especially for patients admitted with cystic fibrosis or immuno-compromising conditions. As other opportunistic Gram-negative bacteria, this pathogen establishes virulence and biofilms through lectin-mediated adhesion. In particular, the superlectin BC2L-C is believed to cross-link human epithelial cells to B. cenocepacia during pulmonary infection.

With the ultimate goal of inhibiting the interactions between the N-terminal of BC2L-C and its target human oligosaccharides, we aim to design glycomimetic antagonists. Here we report the structural study of the target BC2L-C-N-terminal by X-ray crystallography, followed by the design and synthesis of a modular fucoside library of C- and N-glycomimetics. Lastly, we report the biophysical evaluation of the generated glycomimetics against BC2L-CN_{ter} by techniques such as STD-NMR, SPR, ITC, DSC; resulting in a lead structure with satisfactory affinity and two crystal structures of antagonist/lectin complexes.